

AD-A254 745



2

***Tutorials From the Twenty-third
Annual Precise Time and Time
Interval (PTTI) Applications and
Planning Meeting***

S DTIC
ELECTE
AUG 28 1992
A **D**



**A meeting held at the
Ritz-Carlton Huntington Hotel
Pasadena, California
December 3-5, 1991**

This document has been approved
for public release and sale; its
distribution is unlimited.

82

92-23861



390007 1561K

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

| | | | | |
|--|---|--|---|--|
| 1. AGENCY USE ONLY (Leave blank) | | 2. REPORT DATE September 1992 | 3. REPORT TYPE AND DATES COVERED Technical Report | |
| 4. TITLE AND SUBTITLE Tutorials from the Twenty-third Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting | | | 5. FUNDING NUMBERS | |
| 6. AUTHOR(S) J. Vig, Army Research Laboratory; G. Lutes and R. Logan, Jet Propulsion Laboratory; J. Jespersen, NIST; J. Barnes, Austron, Inc., Lt. C. Fox, U.S. Air Force; G. A. Gifford, Naval Research Laboratory; and S. R. Stein, Timing Solutions Corporation | | | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) (submitted from) US Army Electronics Technology and Devices Laboratory ATTN: SLCET-EF Fort Monmouth, NJ 07703-5601 | | | 8. PERFORMING ORGANIZATION REPORT NUMBER | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Naval Observatory, NASA Goddard Space Flight Center, Space and Naval Warfare Systems Command, Naval Research Laboratory, National Institute of Standard and Technology, Army Electronics Technology and Devices Laboratory, Rome Air Development Center, and USAF Space Command. | | | 10. SPONSORING/MONITORING AGENCY REPORT NUMBER | |
| 11. SUPPLEMENTARY NOTES | | | | |
| 12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited. | | | 12b. DISTRIBUTION CODE | |
| 13. ABSTRACT (Maximum 200 words) The tutorial papers in this document are: "Introduction to Quartz Frequency Standards," J. Vig, Army Research Laboratory; "Tutorial on High Performance Analog Fiber Optic Systems," G. Lutes and R. Logan, Jet Propulsion Laboratory; "Introduction to the Time Domain Characterization of Frequency Standards," J. Jespersen, NIST; "Noise Models for Time and Frequency," J. Barnes, Austron, Inc., "GPS Time Determination and Dissemination," Lt. C. Fox, U.S. Air Force; G. A. Gifford, Naval Research Laboratory; and S. R. Stein, Timing Solutions. | | | | |
| 14. SUBJECT TERMS Oscillator, clock, frequency standard, frequency control, frequency stability, time, timing devices, quartz, quartz crystal, quartz oscillator, atomic clock, atomic frequency standard, frequency distribution, stability, aging, noise, phase noise, Allan variance, fiber optics, GPS, Global Positioning System. | | | 15. NUMBER OF PAGES 154 | |
| | | | 16. PRICE CODE | |
| 17. SECURITY CLASSIFICATION OF REPORT Unclassified | 18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified | 19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified | 20. LIMITATION OF ABSTRACT UL | |

Tutorials From the Twenty-third Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting

DTIC QUALITY INSPECTED 5



| | |
|--------------------|--|
| Accession For | |
| NTIS CRA&I | <input checked="checked" type="checkbox"/> |
| DTIC TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Justification | |
| By | |
| Distribution / | |
| Availability Codes | |
| Dist | Avail. and/or Special |
| A-1 | |

**A meeting held at the Ritz-Carlton Huntington Hotel
Pasadena, California
December 3-5, 1991**

CONTENTS

Tutorials

Chairman: C. Wardrip
Bendix Field Engineering Corporation

| | |
|--|------------|
| Introduction to Quartz Frequency Standards | 1 |
| J. Vig, U.S. Army Electronics Technology and Devices Laboratory | |
| Tutorial on High Performance Analog Fiber Optic Systems | 51 |
| G. Lutes and R. Logan, Jet Propulsion Laboratory | |
| Introduction to the Time Domain Characterization of Frequency Standards . | 83 |
| J. Jespersen, NIST | |
| Noise Models for Time and Frequency | 103 |
| J. Barnes, Austron, Inc. | |
| GPS Time Determination and Dissemination | 109 |
| Lt. C. Fox, U.S. Air Force, G.A. Gifford, Naval Research Laboratory, and S.R. Stein, Timing Solutions Corporation | |

Introduction to Quartz Frequency Standards

John R. Vig
U. S. Army Electronics Technology and Devices Laboratory
Fort Monmouth, NJ 07703-5601

Abstract

More than one billion (i.e., 10^9) quartz crystal oscillators are produced annually for applications ranging from inexpensive watches and clocks to radionavigation and spacecraft tracking systems. The fundamentals of quartz oscillators are reviewed in this paper, with emphasis on quartz frequency standards (as opposed to inexpensive clock oscillators). The subjects discussed include: crystal resonators and oscillators, oscillator types, and the characteristics and limitations of temperature-compensated crystal oscillators (TCXO) and oven-controlled crystal oscillators (OCXO). The oscillator instabilities discussed include: aging, noise, frequency vs. temperature, warmup, acceleration effects, magnetic field effects, atmospheric pressure effects, radiation effects, and interactions among the various effects. Guidelines are provided for oscillator comparison and selection. Discussions of specifications are also included, as are references and suggestions for further reading.

Table of Contents

| | <u>Page</u> |
|--|--------------------|
| I. Introduction | 1 |
| II. Crystal Oscillators | 1 |
| A. Generalized Crystal Oscillator | 1 |
| 1. Oscillator Basics | |
| 2. Crystal Unit Equivalent Circuit | 2 |
| 3. Stability versus Tunability | 5 |
| 4. The Quartz Crystal Unit | 6 |
| B. Oscillator Categories | 13 |
| C. Oscillator Circuit Types | 15 |
| III. Oscillator Instabilities | 17 |
| A. Accuracy, Stability and Precision | 17 |
| B. Aging | 18 |
| C. Noise in Frequency Standards | 19 |
| 1. The Effects of Noise | 19 |
| 2. Noise in Crystal Oscillators | 20 |
| D. Frequency versus Temperature Stability | 22 |
| 1. Static Frequency versus Temperature Stability | 22 |
| 2. Dynamic Frequency versus Temperature Effects | 25 |
| 3. Thermal Hysteresis and Retrace | 27 |
| E. Warm-up | 28 |
| F. Acceleration Effects | 29 |
| G. Magnetic-Field Effects | 33 |
| H. Radiation Effects | 34 |
| I. Other Effects on Stability | 36 |
| J. Interactions Among the Influences on Stability | 37 |
| IV. Oscillator Comparison and Selection | 39 |
| V. Specifications, Standards, Terms, and Definitions | 43 |
| VI. For Further Reading | 44 |
| VII. References | 44 |

Figures

| <u>Figure</u> | <u>Page</u> |
|--|-------------|
| 1. Crystal oscillator - simplified circuit diagram. | 2 |
| 2. Equivalent circuit of a mechanically vibrating system. | 3 |
| 3. Equivalent circuit of crystal unit with load capacitor. | 3 |
| 4. Reactance versus frequency of a crystal unit. | 4 |
| 5. Zero-temperature-coefficient cuts of quartz. | 8 |
| 6. Typical constructions of AT-cut and SC-cut crystal units: (a) two-point mount package; (b) three- and four-point mount package. | 9 |
| 7. Resonator vibration amplitude distribution. | 10 |
| 8. Drive level dependence of frequency. | 11 |
| 9. Drive level dependence of crystal unit resistance. | 11 |
| 10. Modes of motion of a quartz resonator. | 12 |
| 11. Frequency versus temperature characteristics of AT-cut crystals, showing AT and BT-cut plates in Y-bar quartz. | 13 |
| 12. Crystal oscillator categories based on the crystal unit's frequency versus temperature characteristics. | 14 |
| 13. Oscillator circuit types. | 15 |
| 14. Oscillator outputs. | 16 |
| 15. Accuracy, stability and precision examples for a marksman, <i>top</i> , and for a frequency source, <i>bottom</i> . | 17 |
| 16. Computer-simulated typical aging behaviors; where $A(t)$ and $B(t)$ are logarithmic functions with different coefficients. | 19 |
| 17. Low-Noise SAW and BAW multiplied to 10 GHz (in a nonvibrating environment). | 21 |

| | | |
|-----|--|----|
| 18. | Low-Noise SAW and BAW multiplied to 10 GHz (in a vibrating environment). | 22 |
| 19. | Wristwatch accuracy as it is affected by temperature. | 23 |
| 20. | Effects of harmonics on f vs. T . | 24 |
| 21. | Activity dips in the frequency versus temperature and resistance versus temperature characteristics, with and without C_L . | 25 |
| 22. | Warm-up characteristics of AT-cut and SC-cut crystal oscillators (OCXOs). | 26 |
| 23. | Temperature-compensated crystal oscillators (TCXO) thermal hysteresis showing that the f vs. T characteristic upon increasing temperature differs from the characteristic upon decreasing temperature. | 27 |
| 24. | Oven-controlled crystal oscillator (OCXO) retrace, showing that upon restarting the oscillator after a 14 day off-period, the frequency was about 7×10^{-9} lower than it was just before turn-off, and that the aging rate had increased significantly upon the restart. | 28 |
| 25. | 2-g tipover test (Δf vs. attitude about three axes). | 30 |
| 26. | Vibration-induced "sidebands" (i.e., spectral lines). | 30 |
| 27. | Resonance in the acceleration sensitivity vs. vibration frequency characteristic. | 31 |
| 28. | Random-vibration-induced phase-noise degradation. | 31 |
| 29. | Coherent radar probability of detection as a function of phase noise. | 33 |
| 30. | The effect of a shock at $t = t_i$ on oscillator frequency. | 34 |
| 31. | Crystal oscillator's response to a pulse of ionizing radiation: f_0 = original preirradiation frequency, Δf_{ss} = steady-state frequency offset (0.2 hours to 24 hours after exposure), f_t = instantaneous frequency at time t . | 35 |
| 32. | Change in compensating frequency versus temperature due to C_L change. | 38 |
| 33. | Temperature-compensated crystal oscillator (TCXO) trim effect. | 38 |

| | | |
|-----|---|----|
| 34. | Relationship between accuracy and power requirements (XO = simple crystal oscillator; TCXO = temperature-compensated crystal oscillator; OCXO = oven-controlled crystal oscillator; Rb = rubidium frequency standard; Cs = cesium beam frequency standard). | 39 |
| 35. | Stability as a function of averaging time comparison of frequency standards. | 41 |
| 36. | Phase instability comparison of frequency standards. | 42 |

Tables

| <u>Table</u> | | <u>Page</u> |
|--------------|---|-------------|
| Table 1. | Salient characteristic comparison of frequency standards. | 40 |
| Table 2. | Weaknesses and wear-out mechanisms comparison of frequency standards. | 42 |

I. Introduction

More than one billion (i.e., 10^9) quartz crystal oscillators are produced annually for applications ranging from inexpensive watches and clocks to radionavigation and spacecraft tracking systems. The fundamentals of quartz oscillators are reviewed in this report, with emphasis on quartz frequency standards (as opposed to inexpensive clock oscillators). The subjects discussed include: crystal resonators and oscillators, oscillator types, and the characteristics and limitations of temperature-compensated crystal oscillators (TCXO) and oven-controlled crystal oscillators (OCXO). The oscillator instabilities discussed include: aging, noise, frequency vs. temperature, warmup, acceleration effects, magnetic field effects, atmospheric pressure effects, radiation effects, and interactions among the various effects. Guidelines are provided for oscillator comparison and selection. Discussions of specifications are also included, as are references and suggestions for further reading.

II. Crystal Oscillators

A. Generalized Crystal Oscillator

1. Oscillator Basics

Figure 1 is a greatly simplified circuit diagram that shows the basic elements of a crystal oscillator [1-3]. The amplifier of a crystal oscillator consists of at least one active device, the necessary biasing networks, and may include other elements for band limiting, impedance matching, and gain control. The feedback network consists of the crystal resonator, and may contain other elements, such as a variable capacitor for tuning.

The frequency of oscillation is determined by the requirement that the closed loop phase shift $= 2n\pi$, where n is an integer, usually 0 or 1. When the oscillator is initially energized, the only signal in the circuit is noise. That component of noise, the frequency of which satisfies the phase condition for oscillation, is propagated around the loop with increasing amplitude. The rate of increase depends on the excess loop gain and on the bandwidth of the crystal network. The amplitude continues to increase until the amplifier gain is reduced, either by the nonlinearities of the active elements (in which case it is *self limiting*) or by an external level-control method.

At steady state, the closed-loop gain $= 1$. If a phase perturbation $\Delta\phi$ occurs, the frequency of oscillation must shift by a Δf in order to maintain the $2n\pi$ phase condition. It can be shown that for a series-resonance oscillator

$$\frac{\Delta f}{f} = - \frac{\Delta\phi}{2Q_L},$$

where Q_L is the loaded Q of the crystal in the network [1]. ("Crystal" and "resonator" are

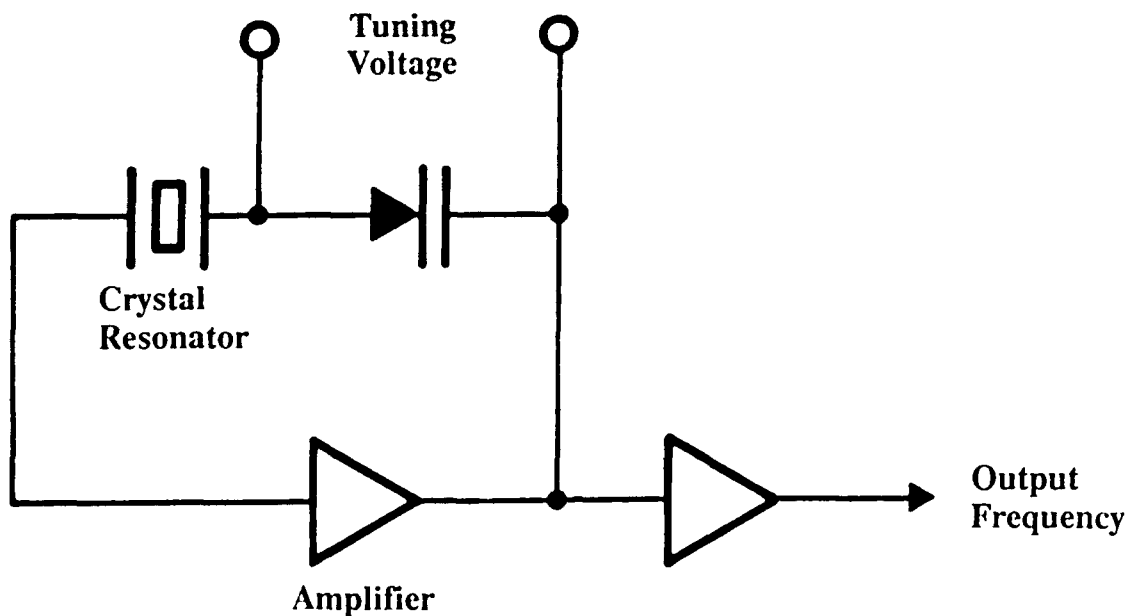


Figure 1. Crystal oscillator - simplified circuit diagram.

often used interchangeably with "crystal unit," although "crystal unit" is the official name. References 3 to 6 contain for further information about crystal units.) Crystal oscillator design information can be found in references 1, 2, 5 and 7. The abbreviation for crystal oscillator is XO.

2. Crystal Unit Equivalent Circuit

A quartz crystal unit is a quartz wafer to which electrodes have been applied, and which is hermetically sealed in a holder structure. (The wafer is often referred to as the "blank".) Although the design and fabrication of crystal units comprise a complex subject, the oscillator designer can treat the crystal unit as a circuit component and just deal with the crystal unit's equivalent circuit.

The mechanically vibrating system and the circuit shown in Figure 2 are "equivalent," because each can be described by the same differential equation [6]. The mass, spring, and damping element (i.e., the dashpot) correspond to the inductor, capacitor and resistor. The driving force corresponds to the voltage, the displacement of the mass to the charge on the capacitor, and the velocity to the current.

A crystal resonator is a mechanically vibrating system that is linked, via the piezoelectric effect, to the electrical world. Figure 3 shows a (simplified) equivalent circuit (of one mode of vibration) of a resonator, together with the circuit symbol for a crystal

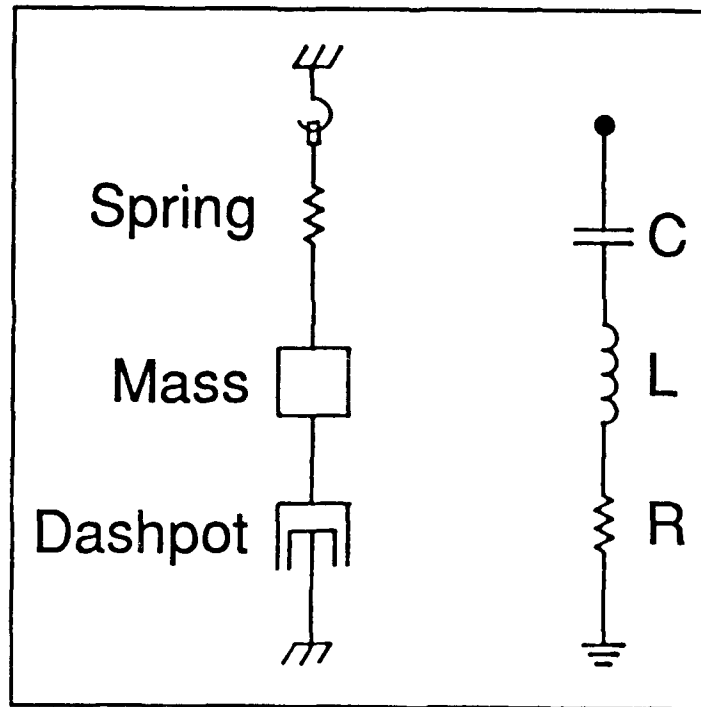


Figure 2. Equivalent circuit of a mechanically vibrating system.

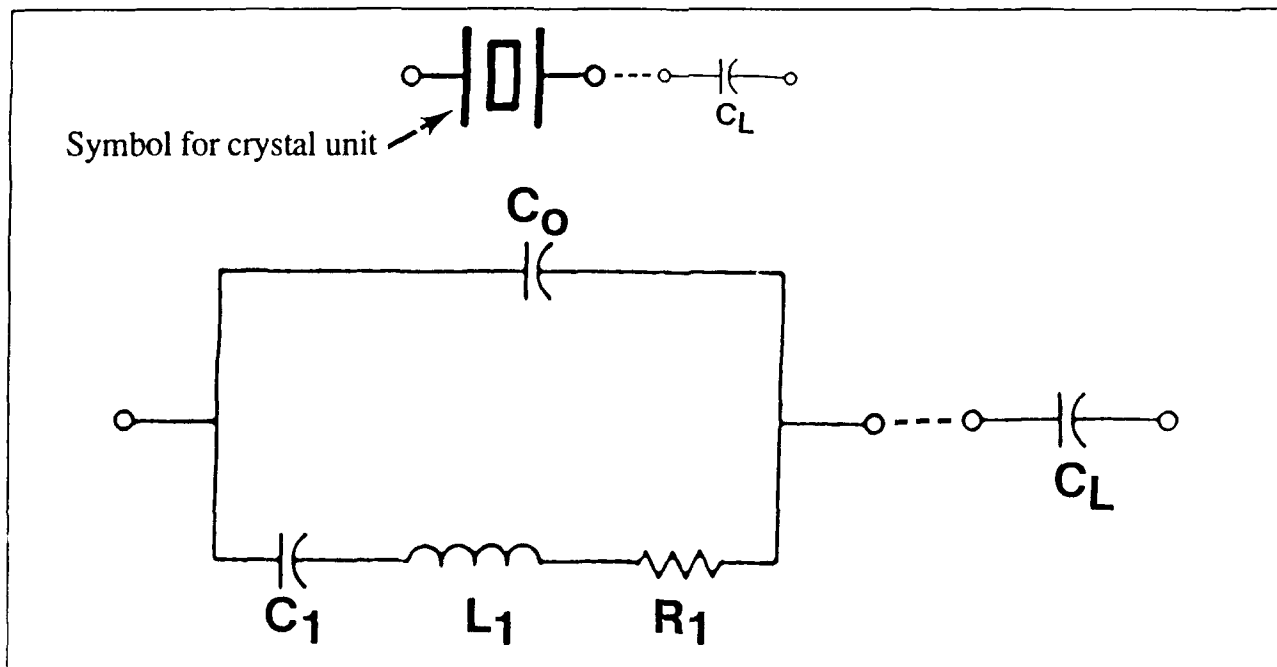


Figure 3. Equivalent circuit of crystal unit with load capacitor.

unit. A load capacitor C_L is shown in series with the crystal. C_0 , called the "shunt" capacitance, is the capacitance due to the electrodes on the crystal plate plus the stray capacitances due to the crystal enclosure. The R_1 , L_1 , C_1 portion of the circuit is the "motional arm" which arises from the mechanical vibrations of the crystal. The C_0 to C_1 ratio is a measure of the interconversion between electrical and mechanical energy stored in the crystal, i.e., of the piezoelectric coupling factor, k . C_0/C_1 increases with the square of the overtone number; the relationship of C_0/C_1 to k and N is $2C_0/C_1 = [\pi N^2/2k]$, where N is the overtone number. When a dc voltage is applied to the electrodes of a resonator, the capacitance ratio C_0/C_1 is also a measure of the ratio of electrical energy stored in the capacitor formed by the electrodes to the energy stored elastically in the crystal due to the lattice strains produced by the piezoelectric effect. Figure 4 shows the reactance versus frequency characteristic of the crystal unit. The C_0/C_1 is also inversely proportional to the antiresonance-resonance frequency separation (i.e., the pole-zero spacing) which is an especially important parameter in filter applications. The slope of the reactance vs. frequency curve near f_s is inversely proportional to C_1 , i.e., $\Delta X/(\Delta f/f) \approx 1/\pi f C_1$ near f_s , where X is the reactance. C_1 is, therefore, a measure of the crystal's "stiffness," i.e., its tunability - see the equation on the next page.

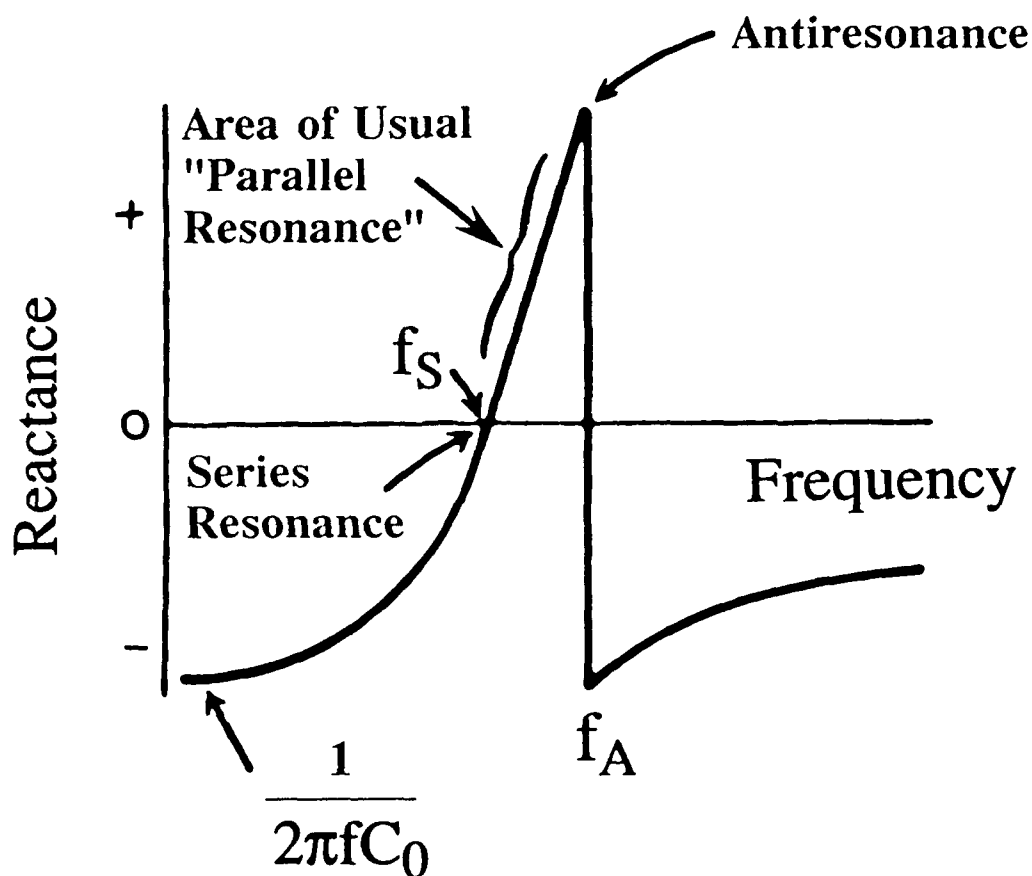


Figure 4. Reactance versus frequency of a crystal unit.

When the load capacitor is connected in series with the crystal, the frequency of operation of the oscillator is increased by a $\Delta f'$, where $\Delta f'$ is given by

$$\frac{\Delta f'}{f} \approx \frac{C_1}{2(C_0 + C_L)}.$$

When an inductor is connected in series with the crystal, the frequency of operation is decreased. The ability to change the frequency of operation by adding or changing a reactance allows for compensation of the frequency versus temperature variations of crystal units in TCXOs, and for tuning the output frequency of voltage controlled crystal oscillators (VCXO); in both, the frequency is changed by changing the voltage on a varactor.

For the simple RLC circuit of Figure 2, the width of the resonance curve is inversely proportional to the quality factor Q , but in a crystal oscillator, the situation is complicated by the presence of C_0 and by the fact that the operating Q is lower than the resonator Q . For a quartz resonator, $Q = (2\pi f_s C_1 R_1)^{-1}$. References 3,5 and 7 contain further details on the equivalent circuit.

Some of the numerous advantages of a quartz crystal resonator over a tank circuit built from discrete R 's, C 's and L 's are that the crystal is far stiffer and has a far higher Q than what could be built from normal discrete components. For example, a 5 MHz fundamental mode AT-cut crystal may have $C_1 = 0.01$ pF, $L_1 = 0.1$ H, $R_1 = 5 \Omega$, and $Q = 10^6$. A 0.01 pF capacitor is not available, since the leads attached to such a capacitor would alone probably contribute more than 0.01 pF. Similarly a 0.1 H inductor would be physically large, it would need to include a large number of turns, and would need to be superconducting in order to have a $\leq 5\Omega$ resistance.

3. Stability versus Tunability

In most crystal oscillator types, a variable-load capacitor is used to adjust the frequency of oscillation to the desired value. Such oscillators operate at the parallel resonance region of Figure 4, where the reactance versus frequency slope (i.e., the "stiffness") is inversely proportional to C_1 . For maximum frequency stability with respect to reactance (or phase) perturbations in the oscillator circuit, the reactance slope (or phase slope) must be maximum. This requires that the C_1 be minimum. The smaller the C_1 , however, the more difficult it is to tune the oscillator (i.e., the smaller is $\Delta f'$ for a given change in C_L). The highest stability oscillators use crystal units that have a small C_1 (and a high Q). Since C_1 decreases rapidly with overtone number, high-stability oscillators generally use third- or fifth-overtone crystal units. Overtones higher than fifth are rarely used, because R_1 also increases rapidly with overtone number, and some tunability is usually desirable in order to allow setting the oscillator to the desired frequency.

Wide-tuning-range VCXOs use fundamental mode crystal units of large C_1 . Voltage control is used for the following purposes: to frequency or phase lock two oscillators; for frequency modulation; for compensation, as in a TCXO (see below); and for calibration (i.e., for adjusting the frequency to compensate for aging). Whereas a high-stability, ovenized 10-MHz VCXO may have a frequency adjustment range of $\pm 5 \times 10^{-7}$ and an aging rate of 2×10^{-8} per year, a wide-tuning-range 10-MHz VCXO may have a tuning range of ± 50 parts per million (ppm) and an aging rate of 2 ppm per year.

In general, making an oscillator tunable over a wide frequency range degrades its stability because making an oscillator susceptible to intentional tuning also makes it susceptible to factors that result in unintentional tuning. For example, if an oven-controlled crystal oscillator (OCXO) is designed to have a stability of 1×10^{-12} for a particular averaging time and a tunability of 1×10^{-7} , then the crystal's load reactance must be stable to 1×10^{-5} for that averaging time. Achieving such load-reactance stability is difficult because the load-reactance is affected by stray capacitances and inductances, by the stability of the varactor's capacitance versus voltage characteristic, and by the stability of the voltage on the varactor. Moreover, the 1×10^{-5} load-reactance stability must be maintained not only under benign conditions, but also under changing environmental conditions (temperature, vibration, radiation, etc.). Therefore, the wider the tuning range of an oscillator, the more difficult it is to maintain a high stability.

4. The Quartz Crystal Unit

A quartz crystal unit's high Q and high stiffness make it the primary frequency and frequency-stability determining element in a crystal oscillator. The Q values of crystal units are much higher than those attainable with other circuit elements. In general-purpose crystal units, Q s are generally in the range of 10^4 to 10^6 . A high-stability 5-MHz crystal unit's Q is typically in the range of two to three million. The intrinsic Q , limited by internal losses in the crystal, has been determined experimentally to be inversely proportional to frequency (i.e., the Qf product is a constant for a given resonator type). For AT- and SC-cut resonators, the maximum $Qf = 16$ million when f is in MHz.

Quartz (which is a single-crystal form of SiO_2) has been the material of choice for stable resonators since shortly after piezoelectric crystals were first used in oscillators - in 1918. Although many other materials have been explored, none has been found to be better than quartz. Quartz is the only material known that possesses the following combination of properties:

1. it is piezoelectric ("pressure electric"; *piezein* means "to press" in Greek)
2. zero temperature coefficient resonators can be made when the plates are cut along the proper directions with respect to the crystallographic axes of quartz
3. of the zero temperature coefficient cuts, one, the SC-cut (see below), is "stress compensated"

4. it has low intrinsic losses (i.e., quartz resonators can have high Q's)
5. it is easy to process because it is hard but not brittle, and, under normal conditions, it has low solubility in everything except the fluoride etchants
6. it is abundant in nature
7. it is easy to grow in large quantities, at low cost, and with relatively high purity and perfection.

Of the man-grown single crystals, quartz, at more than 2000 tons per year (in 1991), is second only to silicon in quantity grown.

The direct piezoelectric effect was discovered by the Curie brothers in 1880. They showed that when a weight was placed on a quartz crystal, charges appeared on the crystal surface; the magnitude of the charge was proportional to the weight. In 1881, the converse piezoelectric effect was illustrated; when a voltage was applied to the crystal, the crystal deformed due to the lattice strains caused by the effect. The strains reversed when the voltage was reversed.

Of the 32 crystal classes, 20 exhibit the piezoelectric effect (but only a few of these are useful). Piezoelectric crystals lack a center of symmetry. When a force deforms the lattice, the centers of gravity of the positive and negative charges in the crystal can be separated so as to produce surface charges. The piezoelectric effect can provide a coupling between an electrical circuit and the mechanical properties of a crystal. Under the proper conditions, a "good" piezoelectric resonator can stabilize the frequency of an oscillator circuit.

Quartz crystals are highly *anisotropic*, that is, the properties vary greatly with crystallographic direction. For example, when a quartz sphere is etched in hydrofluoric acid, the etching rate is more than 100 times faster along the fastest etching rate direction, the Z-direction, than along the slowest direction, the slow-X-direction. The constants of quartz, such as the thermal expansion coefficient and the temperature coefficients of the elastic constants, also vary with direction. That crystal units can have zero temperature coefficients of frequency is a consequence of the temperature coefficients of the elastic constants ranging from negative to positive values.

The locus of zero-temperature-coefficient cuts in quartz is shown in Figure 5. The X, Y, and Z directions have been chosen to make the description of properties as simple as possible. The Z-axis in Figure 5 is an axis of threefold symmetry in quartz; in other words, the physical properties repeat every 120° as the crystal is rotated about the Z-axis. The cuts usually have two-letter names, where the "T" in the name indicates a temperature-compensated cut; for instance, the AT-cut was the first temperature-compensated cut discovered. The FC, IT, BT, and RT-cuts are other cuts along the zero-temperature coefficient locus. These cuts were studied in the past for some special properties, but are rarely used today. The highest-stability crystal oscillators employ SC-cut or AT-cut crystal units.

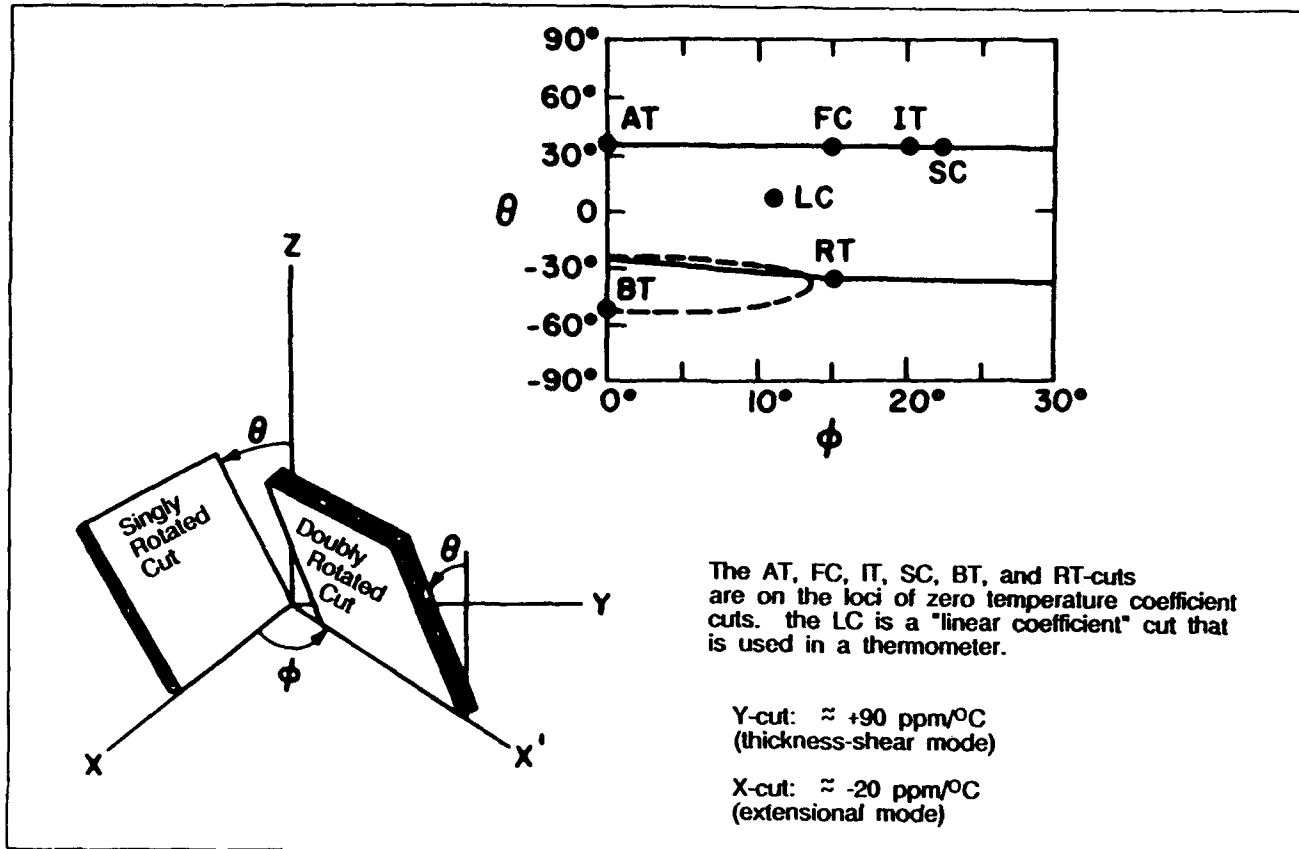


Figure 5. Zero-temperature-coefficient cuts of quartz.

Because the properties of a quartz crystal unit depend strongly on the angles of cut of the crystal plate, in the manufacture of crystal units, the plates are cut from a quartz bar along precisely controlled directions with respect to the crystallographic axes. After shaping to required dimensions, metal electrodes are applied to the wafer, which is mounted in a holder structure [8]. Figure 6 shows the two common types of holder structures used for resonators with frequencies greater than 1 MHz. (The 32-kHz tuning fork resonators used in quartz watches are packaged typically in small tubular enclosures.)

Because quartz is piezoelectric, a voltage applied to the electrodes causes the quartz plate to deform slightly. The amount of deformation due to an alternating voltage depends on how close the frequency of the applied voltage is to a natural mechanical resonance of the crystal. To describe the behavior of a resonator, the differential equations for Newton's laws of motion for a continuum, and for Maxwell's equations, must be solved with the proper electrical and mechanical boundary conditions at the plate surfaces [9]. Because quartz is anisotropic and piezoelectric, with 10 independent linear constants and numerous higher order constants, the equations are complex, and have

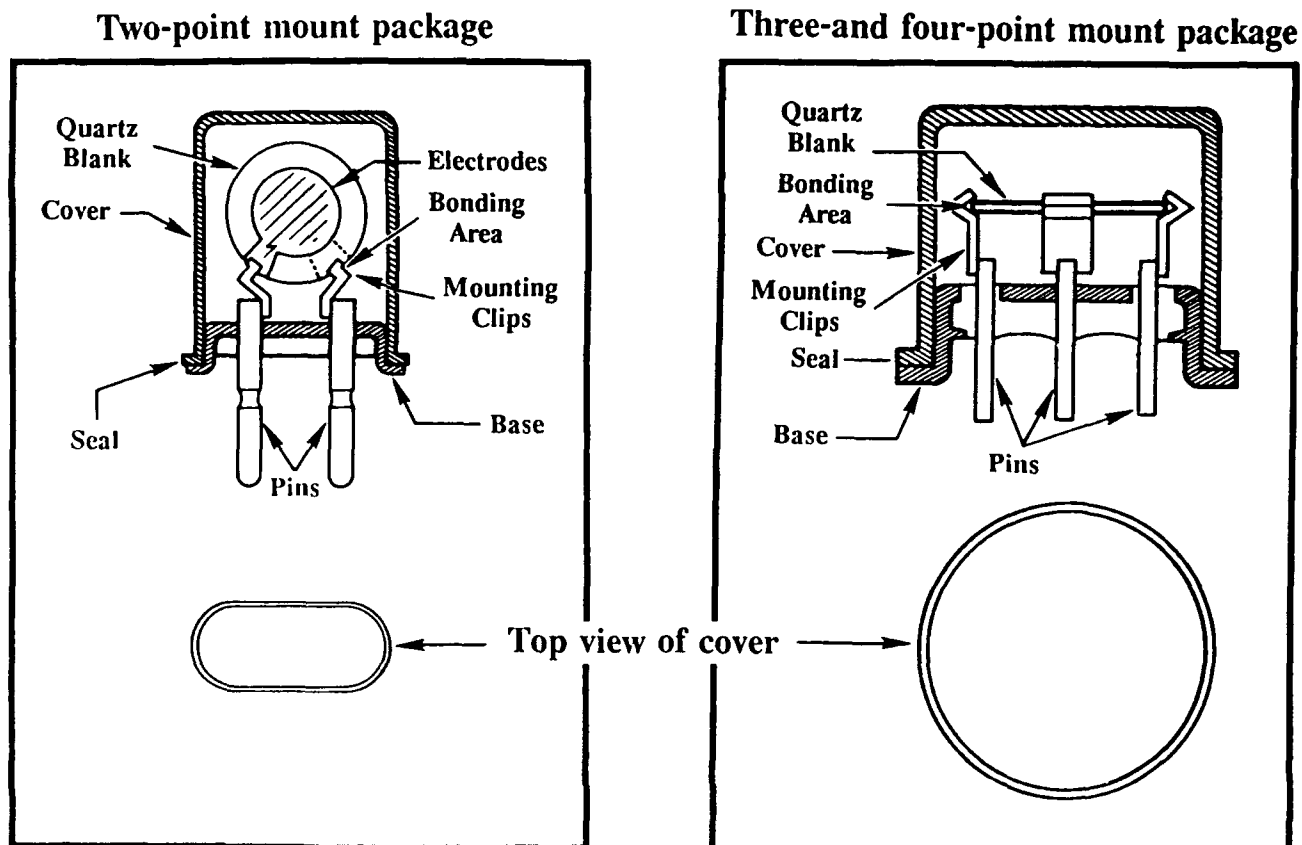


Figure 6. Typical constructions of AT-cut and SC-cut crystal units: (a) two-point mount package; (b) three- and four-point mount package.

never been solved in closed form for physically realizable three-dimensional resonators. Nearly all theoretical works have used approximations. The nonlinear elastic constants, although small, are the source of some of the important instabilities of crystal oscillators; such as the acceleration sensitivity, the thermal-transient effect, and the amplitude-frequency effect, each of which is discussed in this article.

In an ideal resonator, the amplitude of vibration falls off exponentially outside the electrodes, as shown in the lower right portion of Figure 7. In a properly designed resonator, a negligible amount of energy is lost to the mounting and bonding structure, i.e., the edges must be inactive in order for the resonator to be able to possess a high Q . The displacement of a point on the resonator surface is proportional to the drive current. At the typical drive currents used in (e.g., 10 MHz) thickness shear resonators, the peak displacement is on the order of a few atomic spacings.

As the drive level (the current through a crystal) increases, the crystal's amplitude of vibration also increases, and the effects due to the nonlinearities of quartz become

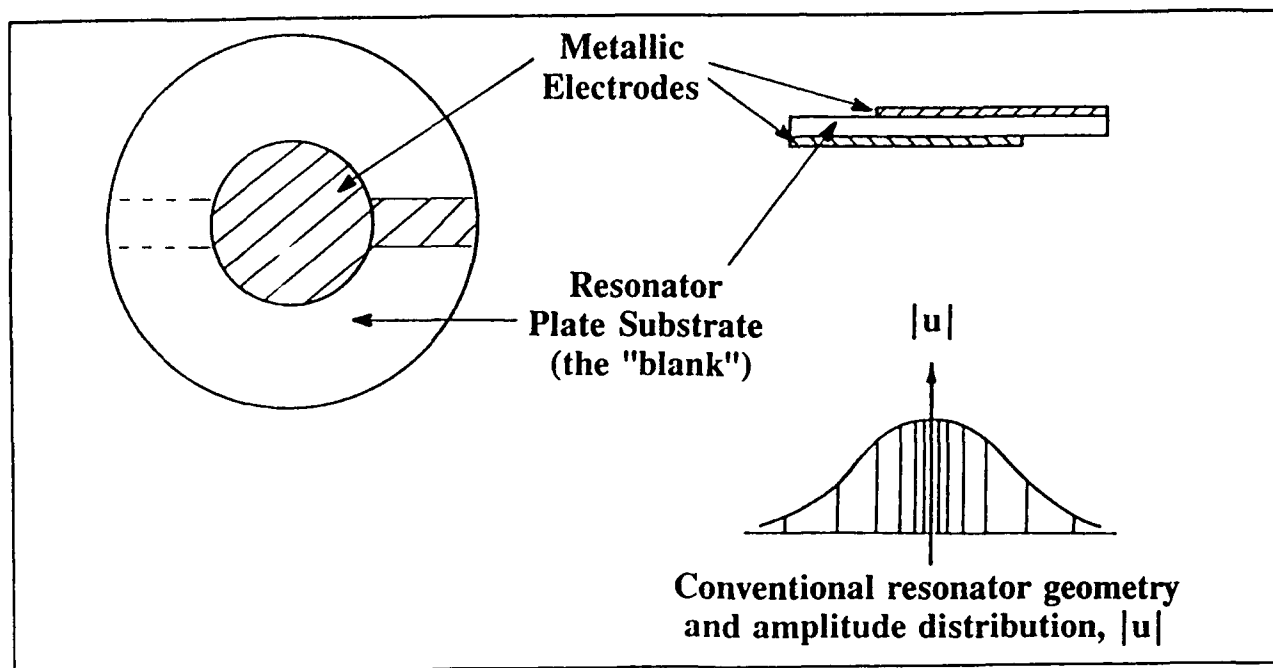


Figure 7. Resonator vibration amplitude distribution.

more pronounced. Among the many properties that depend on the drive level are: resonance frequency, motional resistance R_1 , phase-noise, and frequency vs. temperature anomalies (called *activity dips*), which are discussed in another section of this report. The drive-level dependence of the resonance frequency, called the *amplitude-frequency effect*, is illustrated in Figure 8 [10]. The frequency change with drive level is proportional to the square of the drive current; the coefficient depends on resonator design [11]. Because of the drive-level dependence of frequency, the highest stability oscillators usually contain some form of automatic level control in order to minimize frequency changes due to oscillator circuitry changes. At high drive levels, the nonlinear effects also result in an increase in the resistance [5]. Crystals can also exhibit anomalously high starting resistance when the crystal surfaces possess such imperfections as scratches and particulate contamination. Under such conditions, the resistance at low drive levels can be high enough for an oscillator to be unable to start when power is applied. The drive level dependence of resistance is illustrated in Figure 9. In addition to the nonlinear effects, a high drive level can also cause a frequency change due to a temperature increase caused by the energy dissipation in the active area of the resonator.

Bulk-acoustic-wave quartz resonators are available in the frequency range of about 1 kHz to 500 MHz. Surface-acoustic-wave (SAW) quartz resonators are available in the range of about 150 MHz to 1.5 gigahertz (GHz). To cover the wide range of frequencies, different cuts, vibrating in a variety of modes, are used. The bulk-wave modes of motion are shown in Figure 10. The AT-cut and SC-cut crystals vibrate in a thickness-shear

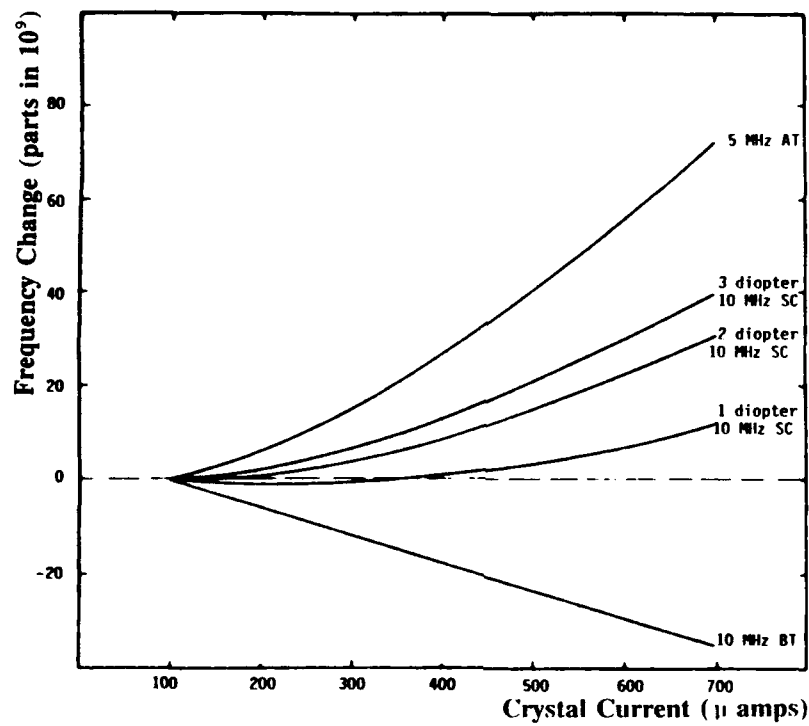


Figure 8. Drive level dependence of frequency.

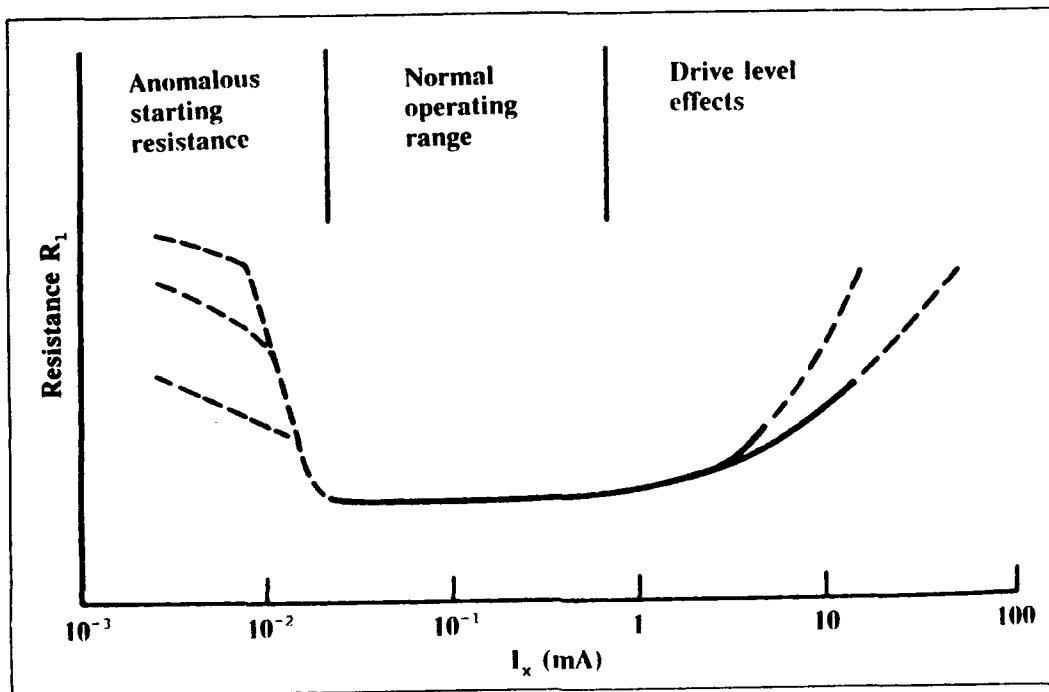


Figure 9. Drive level dependence of crystal unit resistance.

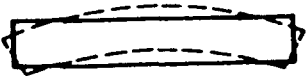
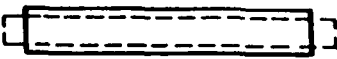
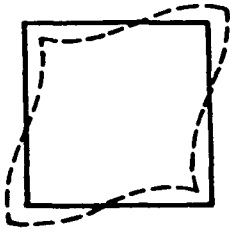


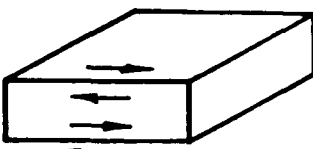
| | | |
|---|---|---|
|  <p data-bbox="211 472 465 514">Flexure Mode</p> |  <p data-bbox="591 472 926 514">Extensional Mode</p> |  <p data-bbox="1025 472 1339 514">Face Shear Mode</p> |
|  <p data-bbox="189 798 492 892">Thickness Shear Mode</p> |  <p data-bbox="581 798 943 892">Fundamental Mode Thickness Shear</p> |  <p data-bbox="1037 798 1326 892">Third Overtone Thickness Shear</p> |

Figure 10. Modes of motion of a quartz resonator.

mode. Although the desired thickness-shear mode usually exhibits the lowest resistance, the mode spectrum of even properly designed crystal units exhibits unwanted modes above the main mode. The unwanted modes, also called "spurious modes" or "spurs," are especially troublesome in filter crystals, in which "energy trapping rules" are employed to maximize the suppression of unwanted modes. These rules specify certain electrode geometry to plate geometry relationships. In oscillator crystals, the unwanted modes may be suppressed sufficiently by providing a large enough plate diameter to electrode diameter ratio, or by contouring (i.e., generating a spherical curvature on one or both sides of the plate).

Above 1 MHz, the AT-cut is commonly used. For high-precision applications, the SC-cut has important advantages over the AT-cut. The AT-cut and SC-cut crystals can be manufactured for fundamental-mode operation up to a frequency of about 200 MHz. (Higher than 1 GHz units have been produced on an experimental basis.) Above 100 MHz, overtone units that operate at a selected harmonic mode of vibration are generally used. Below 1 MHz, tuning forks, X-Y and NT bars (flexure mode), $+5^\circ$ X-cuts (extensional mode), or CT-cut and DT-cut units (face shear mode) can be used. Tuning forks have become the dominant type of low-frequency units due to their small size and low cost. Hundreds of millions of quartz tuning forks are produced annually for quartz watches and other applications.

B. Oscillator Categories

A crystal unit's resonance frequency varies with temperature. Typical frequency vs. temperature (f vs. T) characteristics for crystals used in stable oscillators are shown in Figure 11. The three categories of crystal oscillators, based on the method of dealing with the crystal unit's f vs. T characteristic, are XO, TCXO, and OCXO, (see Figure 12). A simple XO does not contain means for reducing the crystal's f vs. T variation. A typical XO's f vs. T stability may be ± 25 ppm for a temperature range of -55°C to $+85^\circ\text{C}$.

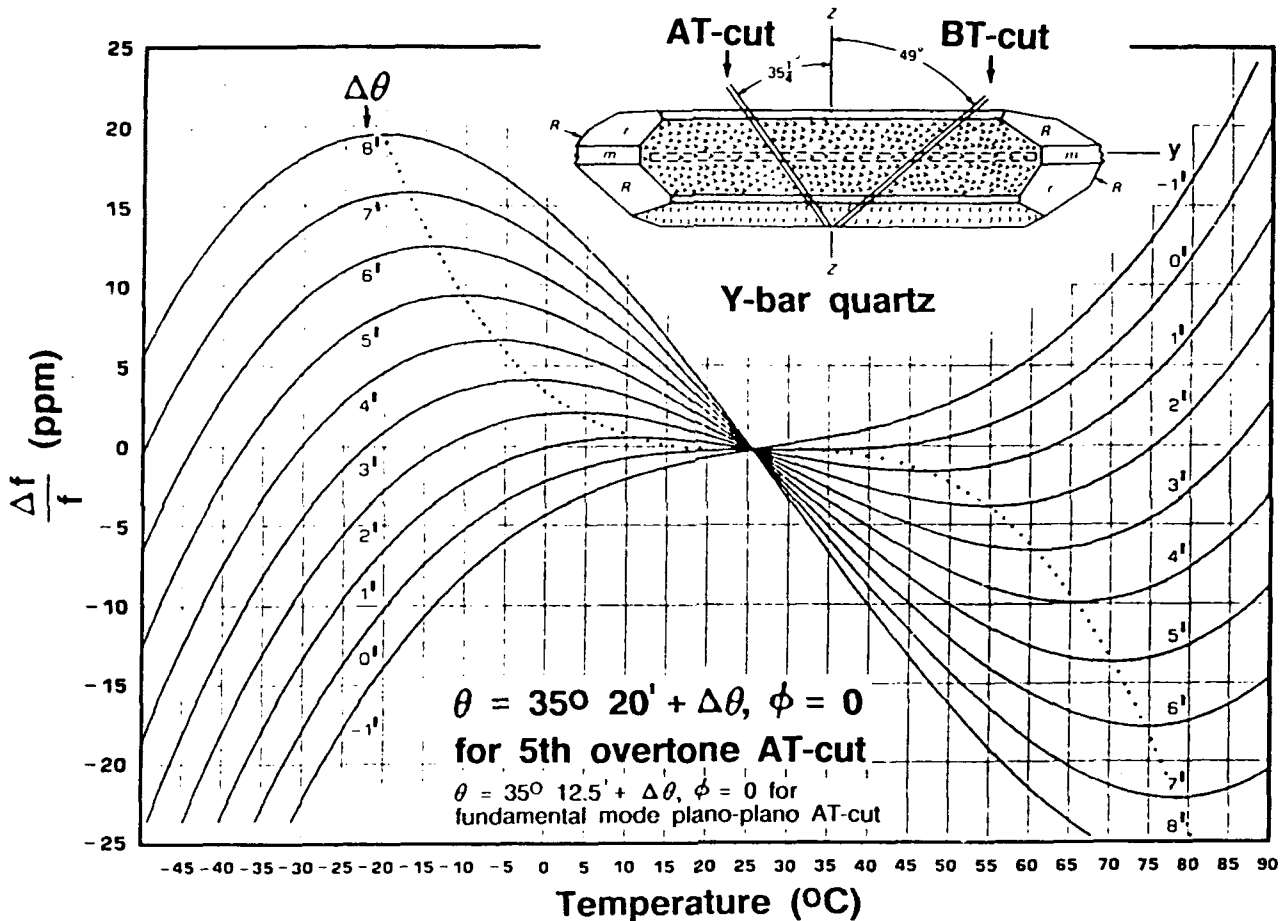


Figure 11. Frequency versus temperature characteristics of AT-cut crystals, showing AT- and BT-cut plates in Y-bar quartz.

In a TCXO, the output signal from a temperature sensor (a *thermistor*) is used to generate a correction voltage that is applied to a voltage-variable reactance (a *varactor*) in the crystal network [12]. The reactance variations produce frequency changes that are equal and opposite to the frequency changes resulting from temperature changes; in other words, the reactance variations compensate for the crystal's f vs. T variations. Analog TCXOs can provide about a 20-fold improvement over the crystal's f vs. T

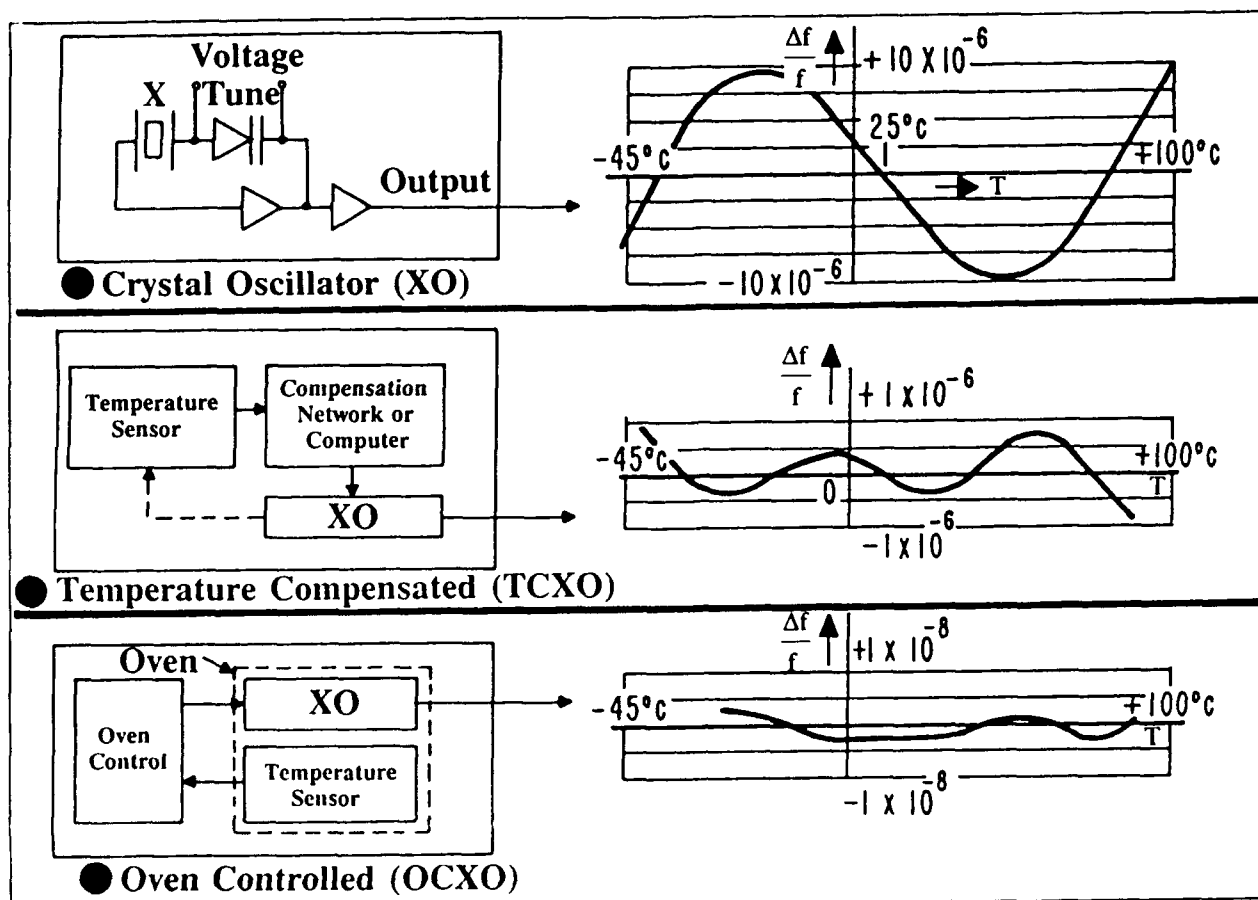


Figure 12. Crystal oscillator categories based on the crystal unit's frequency versus temperature characteristic.

variation. A good TCXO may have an f vs. T stability of ± 1 ppm for a temperature range of -55°C to $+85^\circ\text{C}$.

In an OCXO, the crystal unit and other temperature sensitive components of the oscillator circuit are maintained at a constant temperature in an oven [12]. The crystal is manufactured to have an f vs. T characteristic which has zero slope at the oven temperature. To permit the maintenance of a stable oven temperature throughout the OCXO's temperature range (without an internal cooling means), the oven temperature is selected to be above the maximum operating temperature of the OCXO. OCXOs can provide more than a 1000-fold improvement over the crystal's f vs. T variation. A good OCXO may have an f vs. T stability of better than $\pm 5 \times 10^{-9}$ for a temperature range of -55°C to $+85^\circ\text{C}$. OCXOs require more power, are larger, and cost more than TCXOs.

A special case of a compensated oscillator is the microcomputer-compensated crystal oscillator (MCXO) [13]. The MCXO overcomes the two major factors that limit the

stabilities achievable with TCXOs: thermometry and the stability of the crystal unit. Instead of a thermometer that is external to the crystal unit, such as a thermistor, the MCXO uses a much more accurate "self-temperature sensing" method. Two modes of the crystal are excited simultaneously in a dual-mode oscillator. The two modes are combined such that the resulting beat frequency is a monotonic (and nearly linear) function of temperature. The crystal thereby senses its own temperature. To reduce the f vs. T variations, the MCXO uses digital compensation techniques: pulse deletion in one implementation, and direct digital synthesis of a compensating frequency in another. The frequency of the crystal is not "pulled," which allows the use of high-stability (small C_1) SC-cut crystal units. A typical MCXO may have an f vs. T stability of $\pm 2 \times 10^{-8}$ for a temperature range of -55°C to $+85^\circ\text{C}$.

C. Oscillator Circuit Types

Of the numerous oscillator circuit types, three of the more commonly discussed ones, the Pierce, the Colpitts, and the Clapp, consist of the same circuit except that the rf ground points are at different locations, as shown in Figure 13. The Butler and modified Butler are also similar to each other; in each, the emitter current is the crystal current. The gate oscillator is a Pierce-type that uses a logic gate plus a resistor in place of the transistor in the Pierce oscillator. (Some gate oscillators use more than one gate.)

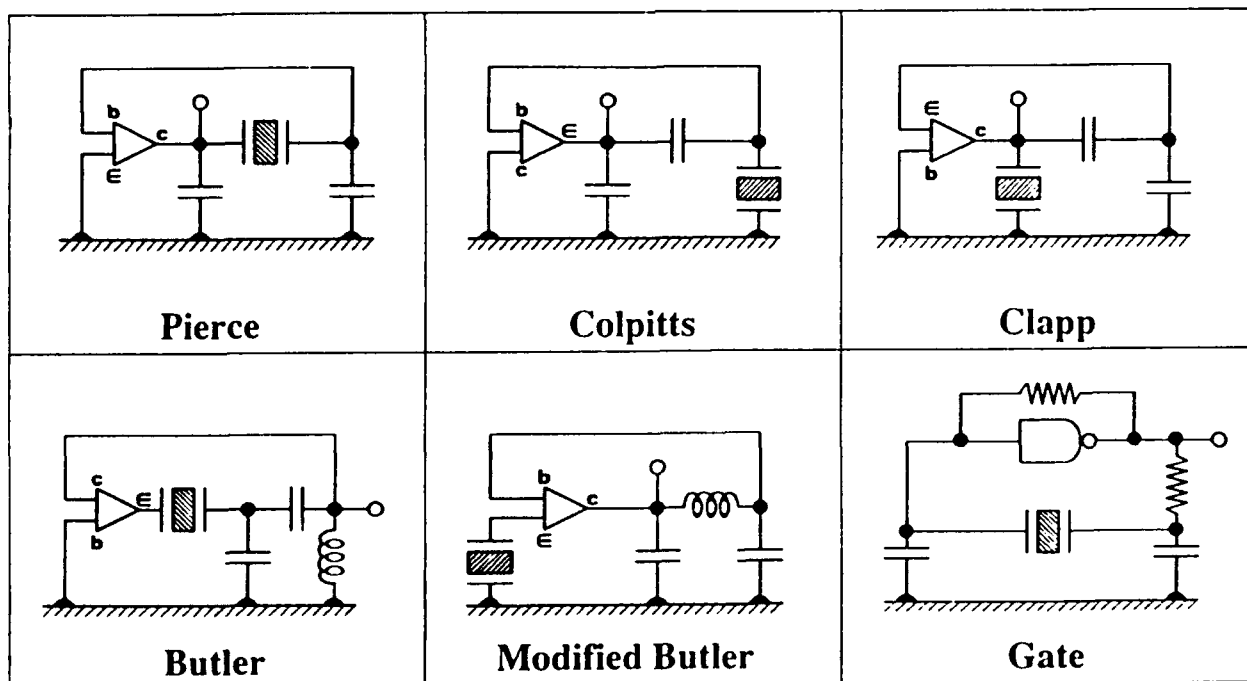


Figure 13. Oscillator circuit types.

Information on designing crystal oscillators can be found in references 2, 5, and 6. The choice of oscillator circuit type depends on factors such as the desired frequency stability, input voltage and power, output power and waveform, tunability, design complexity, cost, and the crystal unit's characteristics.

In the Pierce family, the ground point location has a profound effect on the performance. The Pierce configuration is generally superior to the others, e.g., with respect to the effects of stray reactances and biasing resistors, which appear mostly across the capacitors in the circuit rather than the crystal unit. It is one of the most widely used circuits for high stability oscillators. In the Colpitts configuration, a larger part of the strays appears across the crystal, and the biasing resistors are also across the crystal, which can degrade performance. The Clapp is seldom used because, since the collector is tied directly to the crystal, it is difficult to apply a dc voltage to the collector without introducing losses or spurious oscillations. (See the references for more details.)

The Pierce family usually operates at "parallel resonance" (see Figure 4), although it can be designed to operate at series resonance by connecting an inductor in series with the crystal. The Butler family usually operates at (or near) series resonance. The Pierce can be designed to operate with the crystal current above or below the emitter current. Gate oscillators are common in digital systems when high stability is not a major consideration.

Most users require a sine wave, or a TTL-compatible, or a CMOS-compatible, or an ECL-compatible output. The latter three can be simply generated from a sine wave. The four output types are illustrated in Figure 12, with the dashed lines representing the supply voltage inputs, and the bold solid lines, the outputs. (There is no "standard" input voltage for sine wave oscillators, and the input voltage for CMOS typically ranges from 5V to 15V.)

II. Oscillator Instabilities

A. Accuracy, Stability and Precision

Oscillators exhibit a variety of instabilities. These include aging, noise, and frequency changes with temperature, acceleration, ionizing radiation, power supply voltage, etc. The terms *accuracy*, *stability*, and *precision* are often used in describing an oscillator's quality with respect to its instabilities. Figure 13 illustrates the meanings of these terms for a marksman and for a frequency source. (For the marksman, each bullet hole's distance to the center of the target is the "measurement.") *Accuracy* is the extent to which a given measurement, or the average of a set of measurements for one sample, agrees with the definition of the quantity being measured. It is the degree of "correctness" of a quantity. Atomic frequency standards have varying degrees of accuracy. The International System (SI) of units for time and frequency (second and Hz, respectively) are obtained in laboratories using very accurate frequency standards called

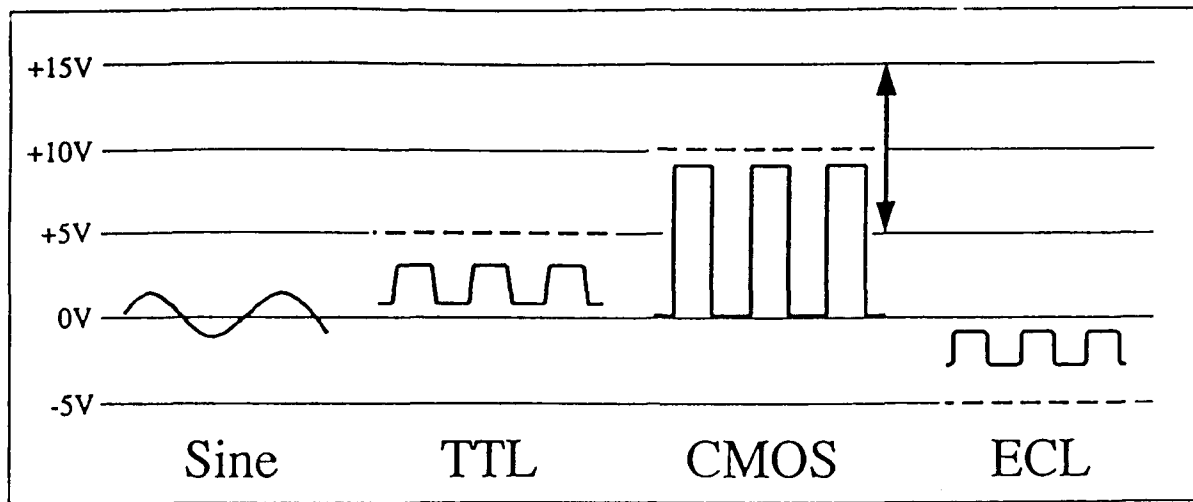


Figure 14. Oscillator outputs.

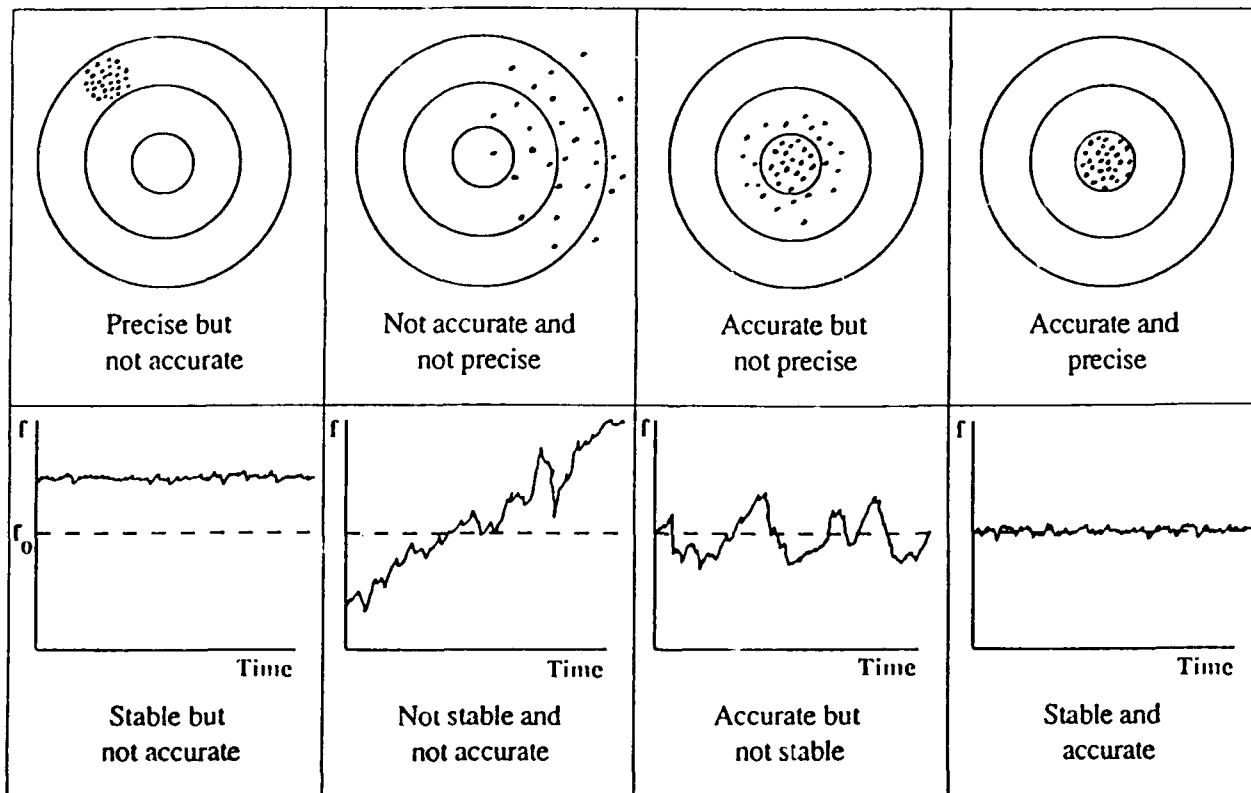


Figure 15. Accuracy, stability and precision examples for a marksman, *top*, and for a frequency source, *bottom*.

primary standards. A primary standard operates at a frequency calculable in terms of the SI definition of the second: "the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium atom 133" [12]. *Reproducibility* is the ability of a single frequency standard to produce the same frequency, without adjustment, each time it is put into operation. From the user's point of view, once a frequency standard is calibrated, reproducibility confers the same advantages as accuracy. *Stability* describes the amount something changes as a function of parameters such as time, temperature, shock, and the like. *Precision* is the extent to which a given set of measurements of one sample agrees with the mean of the set. (A related meaning of the term is used as a descriptor of the quality of an instrument, as in a "precision instrument." In that context, the meaning is usually defined as accurate and precise, although a precision instrument can also be inaccurate and precise, in which case the instrument needs to be calibrated.)

B. Aging

"Aging" and "drift" have occasionally been used interchangeably in the literature. However, in 1990, recognizing the "need for common terminology for the unambiguous specification and description of frequency and time standard systems," the International Radio Consultative Committee (CCIR) adopted a glossary of terms and definitions [13]. According to this glossary, *aging* is "the systematic change in frequency with time due to internal changes in the oscillator," and *drift* is "the systematic change in frequency with time of an oscillator." *Drift* is due to aging plus changes in the environment and other factors external to the oscillator. Aging is what one denotes in a specification document and what one measures during oscillator evaluation. Drift is what one observes in an application. For example, the drift of an oscillator in a spacecraft might be due to (the algebraic sum of) aging and frequency changes due to radiation, temperature changes in the spacecraft, and power supply changes.

Aging can be positive or negative [14]. Occasionally, a reversal in aging direction is observed. At a constant temperature, aging usually has an approximately logarithmic dependence on time. Typical (computer-simulated) aging behaviors are illustrated in Figure 14, where $A(t)$ is a logarithmic function and $B(t)$ is the same function but with different coefficients. The curve showing the reversal is the sum of the other two curves. A reversal indicates the presence of at least two aging mechanisms. The aging rate of an oscillator is highest when it is first turned on. When the temperature of a crystal unit is changed (e.g., when an OCXO is turned off and turned on at a later time), a new aging cycle starts. (See the section concerning hysteresis and retrace below for additional discussion of the effects of temperature cycling).

The primary causes of crystal oscillator aging are mass transfer to or from the resonator's surfaces due to adsorption or desorption of contamination, stress relief in the mounting structure of the crystal, and, possibly, changes in the quartz material. Because the frequency of a thickness-shear crystal unit, such as an AT-cut or an SC-cut, is

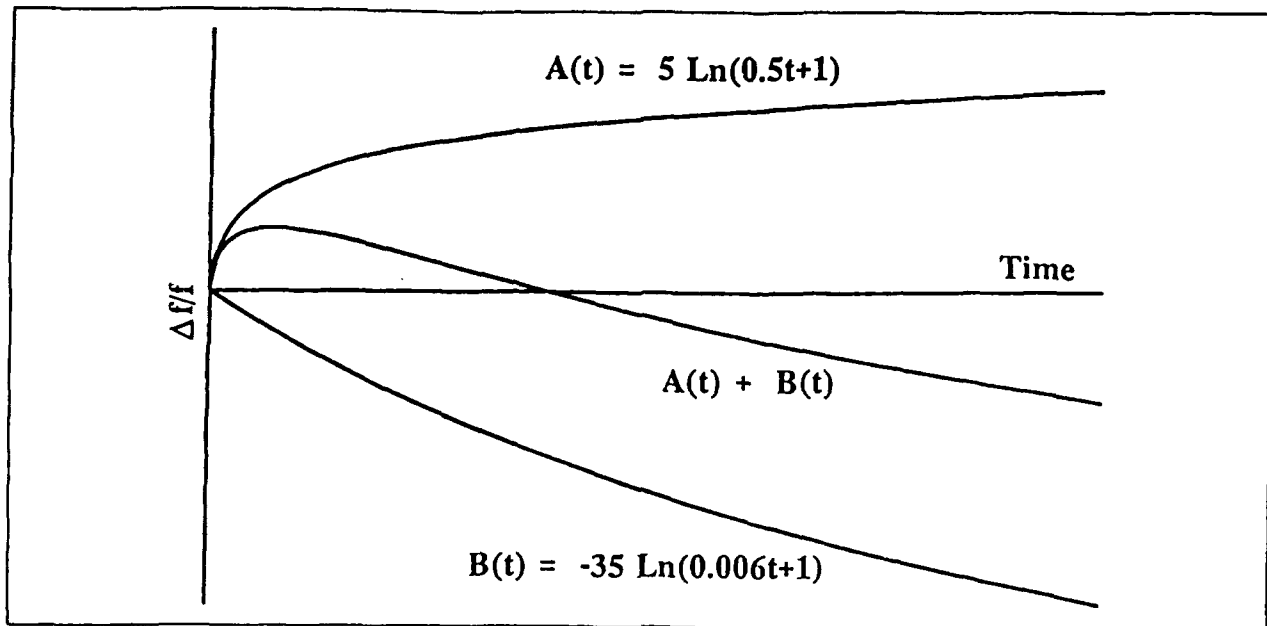


Figure 16. Computer-simulated typical aging behaviors; where $A(t)$ and $B(t)$ are logarithmic functions with different coefficients.

inversely proportional to the thickness of the crystal plate, and because a typical 5-MHz plate is on the order of 1 million atomic layers thick, the adsorption or desorption of contamination equivalent to the mass of one atomic layer of quartz changes the frequency by about 1 ppm. Therefore, in order to achieve low aging, crystal units must be fabricated and hermetically sealed in an ultraclean, ultra-high-vacuum environment. As of 1991, the aging rates of typical commercially available XOs range from 5 ppm to 10 ppm per year for an inexpensive XO, to 0.5 ppm to 2 ppm per year for a TCXO, and to 0.05 ppm to 0.1 ppm per year for an OCXO. The highest precision OCXOs can age less than 0.01 ppm per year.

C. Noise in Frequency Standards

1. The Effects of Noise

Sometimes the suitability of oscillators for an application is limited by deterministic phenomena. In other instances, stochastic (random) processes establish the performance limitations. Except for vibration, the short-term instabilities almost always result from noise. Long-term performance of quartz and rubidium standards is limited by the temperature sensitivity and the aging, but the long-term performance of cesium and some hydrogen standards is limited by random processes.

Noise can have numerous adverse effects on system performance. Among these effects are the following: (1) it limits the ability to determine the current state and the predictability of precision oscillators (e.g., the noise of an oscillator produces time prediction errors of $\sim \tau \sigma_y(\tau)$ for prediction intervals of τ); (2) it limits synchronization and syntonization accuracies; (3) it can limit a receiver's useful dynamic range, channel spacing, and selectivity; (4) it can cause bit errors in digital communications systems; (5) it can cause loss of lock, and limit acquisition and reacquisition capability in phase-locked-loop systems; and (6) it can limit radar performance, especially Doppler radar.

It is important to have appropriate statistical measures to characterize the random component of oscillator instability. The subject of noise characterization has been reviewed [4,17] and is also the subject of an IEEE standard [18].

2. Noise in Crystal Oscillators

Although the causes of noise in crystal oscillators are not fully understood, several causes of short-term instabilities have been identified. Temperature fluctuations can cause short-term instabilities via thermal-transient effects (see the section below concerning dynamic f vs. T effects), and via activity dips at the oven set point in OCXOs. Other causes include Johnson noise in the crystal unit, random vibration (see the section below concerning acceleration effects in crystal oscillators), noise in the oscillator circuitry (both the active and passive components can be significant noise sources), and fluctuations at various interfaces on the resonator (e.g., in the number of molecules adsorbed on the resonator's surface).

In a properly designed oscillator, the resonator is the primary noise source close to the carrier and the oscillator circuitry is the primary source far from the carrier. The noise close to the carrier (i.e., within the bandwidth of the resonator) has a strong inverse relationship with resonator Q , such that $S_y(f) \propto 1/Q^4$. In the time domain, $\sigma_y(\tau) \approx (2 \times 10^{-7})/Q$ at the noise floor. In the frequency domain, the noise floor is limited by Johnson noise, the noise power of which is $kT = -174$ dBm/Hz at 290°K. A higher signal (i.e., a higher resonator drive current) will improve the noise floor but not the close-in noise. In fact, for reasons that are not understood fully, above a certain point, higher drive levels generally degrade the close-in noise. For example, the maximum "safe" drive level is about 100 μ A for a 5-MHz fifth overtone AT-cut resonator with $Q \approx 2.5$ million. The safe drive current can be substantially higher for high-frequency SC-cut resonators. For example, $\mathcal{L}(f) = -180$ dBc/Hz has been achieved with 100-MHz fifth overtone SC-cut resonators at drive currents ≈ 10 mA. However, such a noise capability is useful only in a vibration-free environment, for if there is vibration at the offset frequencies of interest, the vibration-induced noise will dominate the quiescent noise of the oscillator (see the section below concerning acceleration effects in crystal oscillators).

When low noise is required in the microwave (or higher) frequency range, SAW oscillators and dielectric resonator oscillators (DROs) are sometimes used. When compared with multiplied-up (bulk-acoustic-wave) quartz oscillators, these oscillators can provide lower noise far from the carrier at the expense of poorer noise close to the carrier, poorer aging, and poorer temperature stability. SAW oscillators and DROs can provide lower noise far from the carrier because these devices can be operated at higher drive levels, thereby providing higher signal-to-noise ratios, and because the devices operate at higher frequencies, thereby reducing the "20 log N" losses due to frequency multiplication by N . $\mathcal{L}(f) = -180$ dBc/Hz noise floors have been achieved with state-of-the-art SAW oscillators [19]. Of course, as is the case for high-frequency bulk-wave oscillators, such noise floors are realizable only in environments that are free of vibrations at the offset frequencies of interest. Figure 17 and 18 show comparisons of state-of-the-art 5 MHz and 100 MHz BAW oscillators and a 500 MHz SAW oscillator, multiplied to 10 GHz. Figure 17 shows the comparison in a quiet environment, and Figure 18 shows it in a vibrating environment.

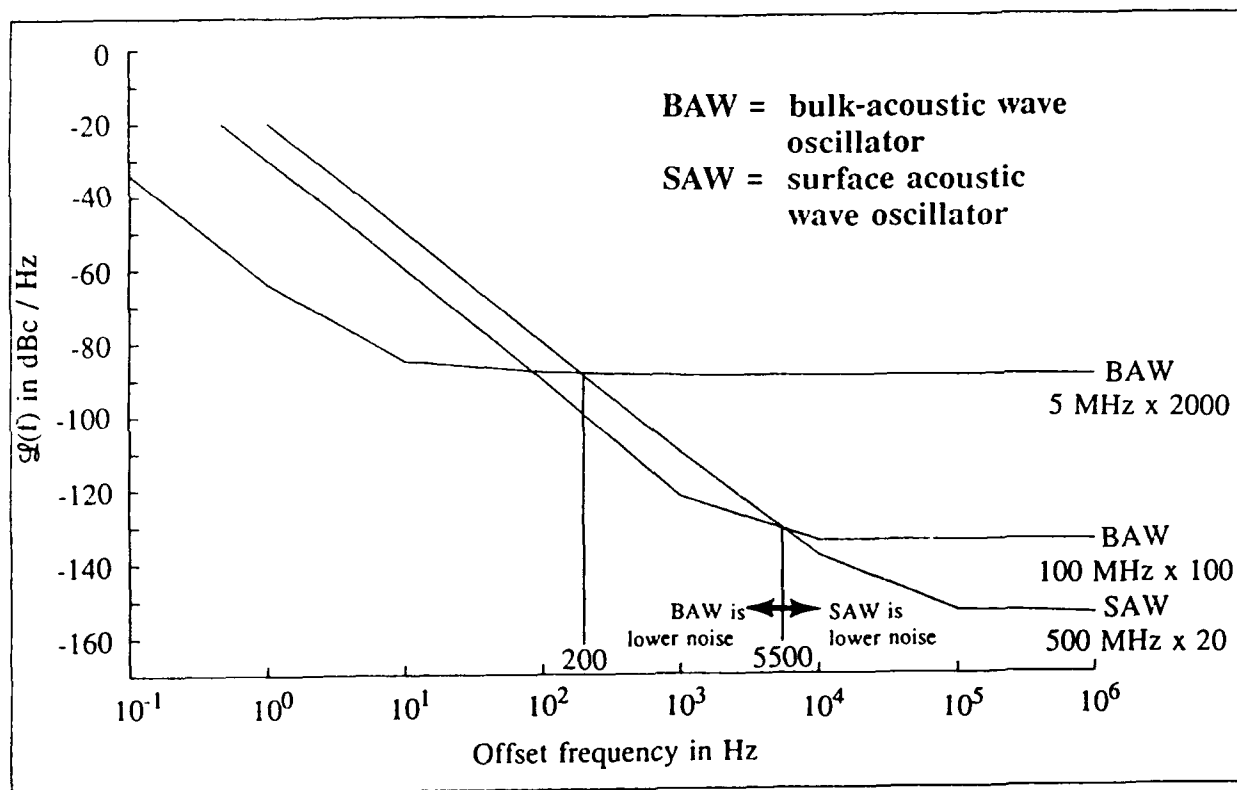


Figure 17. Low-noise SAW and BAW multiplied to 10 GHz (in a nonvibrating environment).

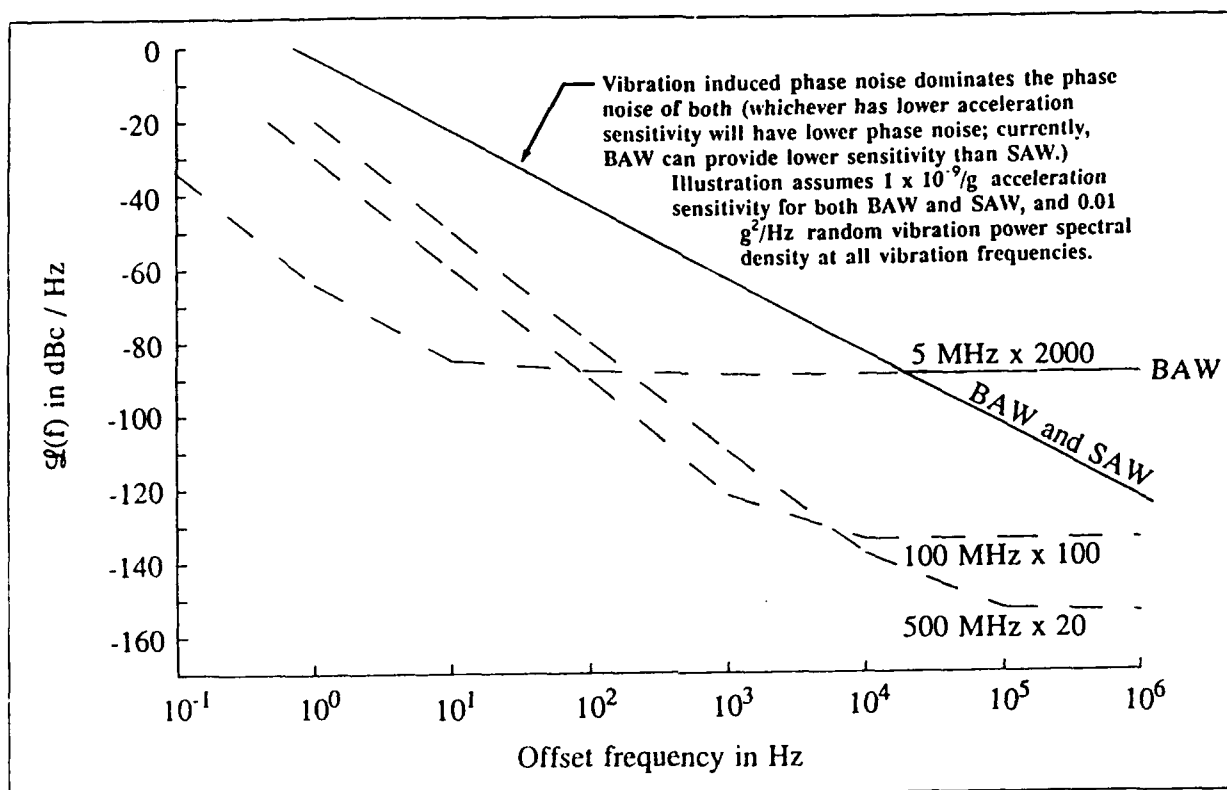


Figure 18. Low-noise SAW and BAW multiplied to 10 GHz (in a vibrating environment).

D. Frequency versus Temperature Stability

1. Static Frequency versus Temperature Stability

As an illustration of the effects that temperature can have on frequency stability, Figure 19 shows the effects of temperature on the accuracy of a typical quartz wristwatch. Near the wrist temperature, the watch can be very accurate because the frequency of the crystal (i.e., the clock rate) changes very little with temperature. However, when the watch is cooled to -55°C or heated to $+100^{\circ}\text{C}$, it loses about 20 seconds per day, because the typical temperature coefficient of the frequency of the tuning fork crystals used in quartz watches is $-0.035 \text{ ppm}/^{\circ}\text{C}^2$.

The static f vs. T characteristics of crystal units are determined primarily by the angles of cut of the crystal plates with respect to the crystallographic axes of quartz [3-5]. "Static" means that the rate of change of temperature is slow enough for the effects of temperature gradients (explained later) to be negligible. As Figure 11 illustrates for the AT-cut, a small change in the angle of cut (seven minutes in the illustration) can significantly change the f vs. T characteristics. The points of zero temperature coefficient,

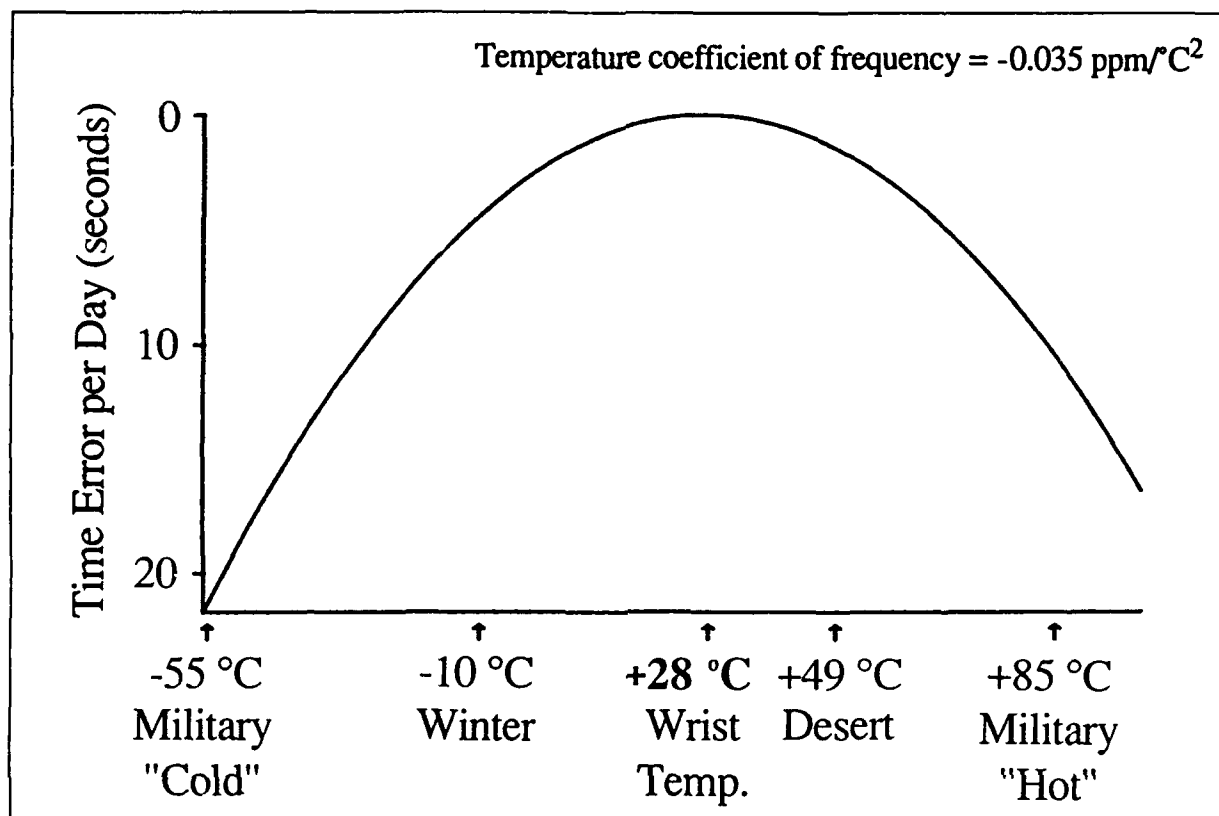


Figure 19. Wristwatch accuracy as it is affected by temperature.

the "turnover points," can be varied over a wide range by varying the angles of cut. The f vs. T characteristics of SC-cut crystals are similar to the curves shown in Figure 11, with the inflection temperature (T_i) shifted to about 95°C .

Other factors that can affect the f vs. T characteristics of crystal units include the overtone [20]; the geometry of the crystal plate; the size, shape, thickness, density and stresses of the electrodes; the drive level; impurities and strains in the quartz material; stresses in the mounting structure; interfering modes; ionizing radiation; the rate of change of temperature (i.e., thermal gradients) [21]; and thermal history. The last two factors are important for understanding the behaviors of OCXOs and TCXOs, and are, therefore, discussed separately.

The effect of harmonics, i.e. "overtones," on f vs. T is illustrated for AT-cut crystals in Figure 20 [20]. This effect is important for understanding the operation of the MCXO. The MCXO contains an SC-cut resonator and a dual mode oscillator that excites both the fundamental mode and the third overtone of the resonator. The difference between the fundamental mode f vs. T and the third overtone f vs. T is due almost exclusively to the

difference between the first order temperature coefficients. Therefore, when the third overtone frequency is subtracted from three times the fundamental mode frequency, the resulting "beat frequency" is a monotonic and nearly linear function of temperature. This beat frequency enables the resonator to sense its own temperature.

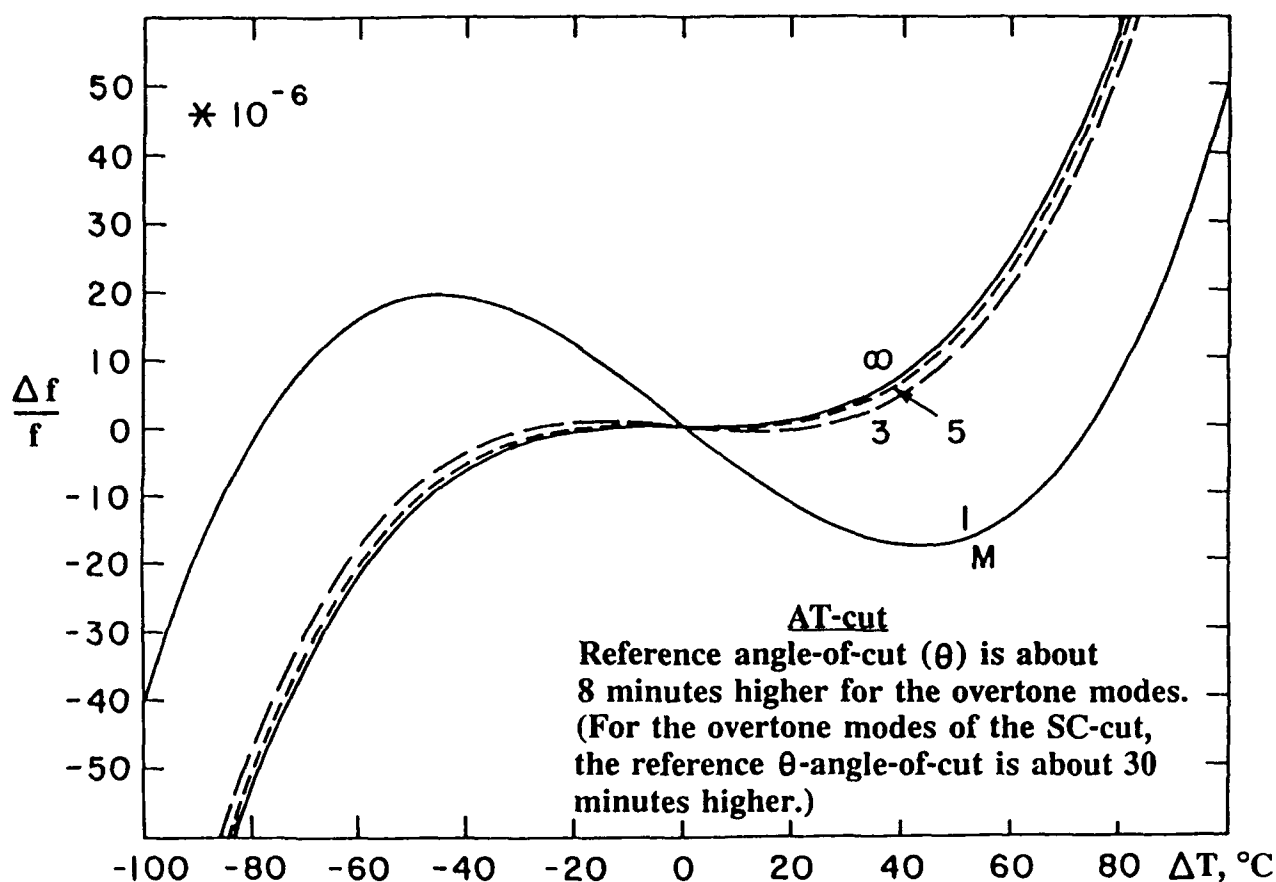


Figure 20. Effects of harmonics on f vs. T .

Interfering modes can cause "activity dips" [22] (see Figure 21). Near the activity dip temperature, anomalies appear in both the f vs. T and resistance (R) vs. T characteristics. Activity dips can be strongly influenced by the crystal's drive level and load reactance. The activity-dip temperature is a function of C_L because the interfering mode usually has a large temperature coefficient and a C_1 that is different from that of the desired mode. Activity dips are troublesome in TCXOs, and also in OCXOs when the dip occurs at the oven temperature. When the resistance increases at the activity dip, and the oscillator's gain margin is insufficient, the oscillation stops. The incidence of activity dips in SC-cut crystals is far lower than in AT-cut crystals.

An important factor that affects the f vs. T characteristics of crystal oscillators is the load capacitor. When a capacitor is connected in series with the crystal, the f vs. T

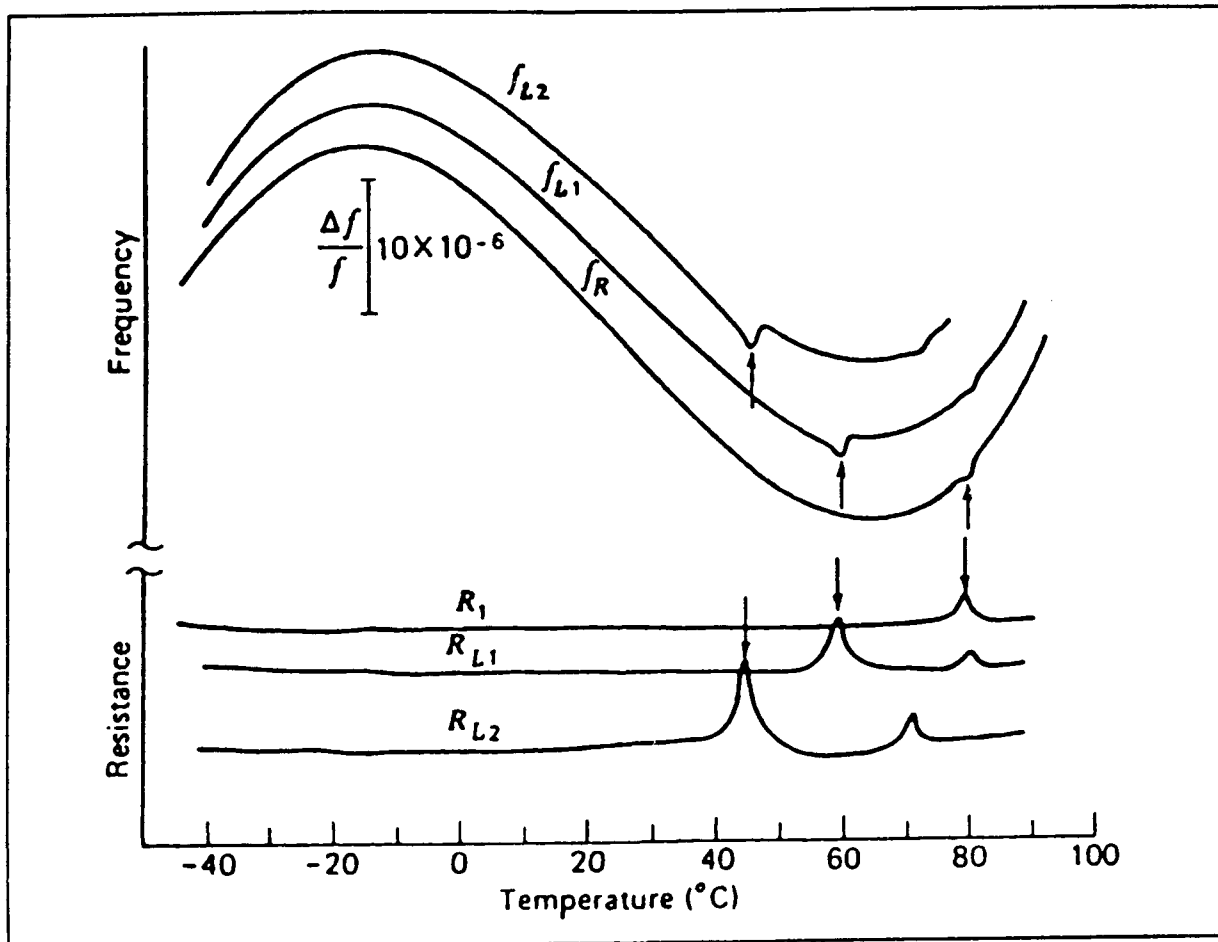


Figure 21. Activity dips in the frequency versus temperature and resistance versus temperature characteristics, with and without C_L .

characteristic of the combination is rotated slightly from that of the crystal alone. The temperature coefficient of the load capacitor can greatly magnify the rotation [23].

The f vs. T of crystals can be described by a polynomial function. A cubic function is usually sufficient to describe the f vs. T of AT-cut and SC-cut crystals to an accuracy of ± 1 ppm. In the MCXO, in order to fit the f vs. T data to $\pm 1 \times 10^{-8}$, a polynomial of at least seventh order is usually necessary [24].

2. Dynamic Frequency versus Temperature Effects

Changing the temperature surrounding a crystal unit produces thermal gradients when, for example, heat flows to or from the active area of the resonator plate through the mounting clips. The static f vs. T characteristic is modified by the thermal-transient

effect. When an OCXO is turned on, there can be a significant thermal-transient effect. Figure 22 shows what happens to the frequency output of two OCXOs, each containing an oven that reaches the equilibrium temperature in six minutes. One oven contains an AT-cut, the other, an SC-cut crystal. Thermal gradients in the AT-cut produce a large frequency undershoot that anneals out several minutes after the oven reaches equilibrium. The SC-cut crystal, being insensitive to such thermal transients, reaches the equilibrium frequency as soon as the oven stabilizes.

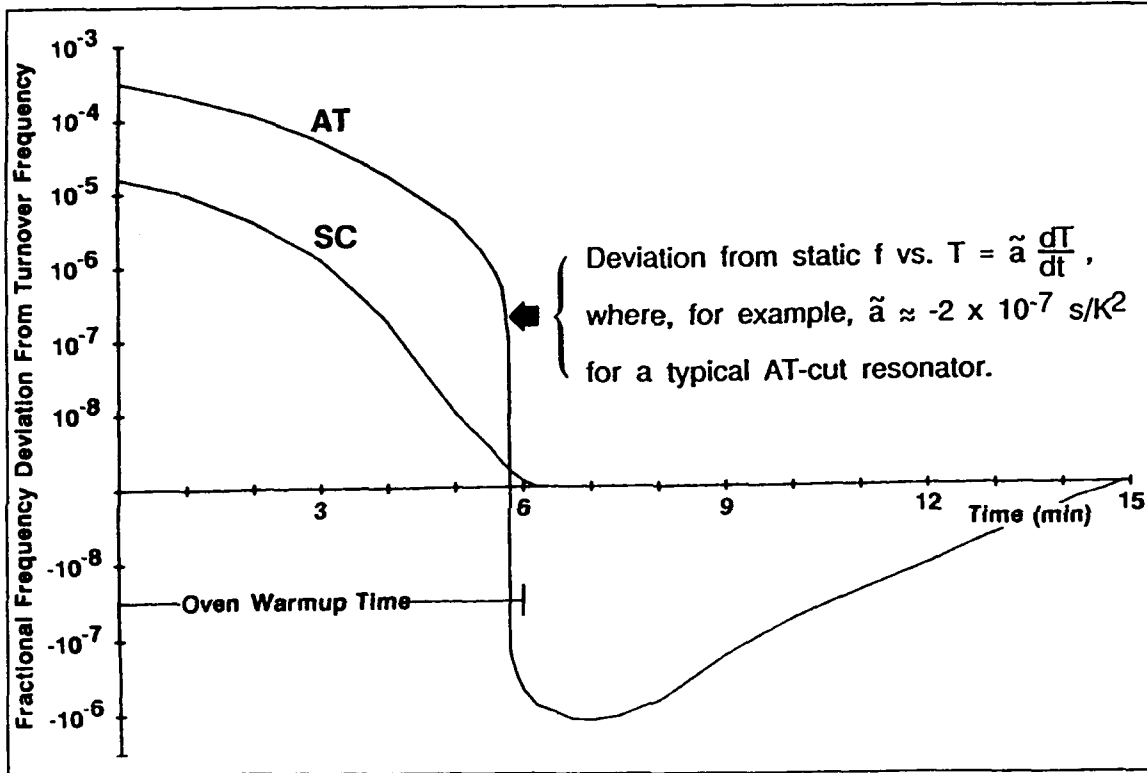


Figure 22. Warm-up characteristics of AT-cut and SC-cut crystal oscillators (OCXOs).

In addition to extending the warmup time of OCXOs, when crystals other than SC-cuts are used, the thermal-transient effect makes it much more difficult to adjust the temperature of OCXO ovens to the desired turnover points, and the OCXO frequencies are much more sensitive to oven-temperature fluctuations [21].

The testing and compensation accuracies of TCXOs are also adversely affected by the thermal-transient effect. As the temperature is changed, the thermal-transient effect distorts the static f vs. T characteristic, which leads to apparent hysteresis [25]. The faster the temperature is changed, the larger is the contribution of the thermal-transient effect to the f vs. T performance.

3. Thermal Hysteresis and Retrace

The f vs. T characteristics of crystal oscillators do not repeat exactly upon temperature cycling [26]. The lack of repeatability in TCXOs, "thermal hysteresis," is illustrated in Figure 23. The lack of repeatability in OCXOs, "retrace," is illustrated in Figure 24. *Hysteresis* is defined [27] as the difference between the up-cycle and the down-cycle f vs. T characteristics, and is quantified by the value of the difference at the temperature where the difference is maximum. Hysteresis is determined during a complete quasistatic temperature cycle between specified temperature limits. *Retrace* is defined as the nonrepeatability of the f vs. T characteristic at a fixed temperature (which is usually the oven temperature of an OCXO) upon on-off cycling an oscillator under specified conditions.

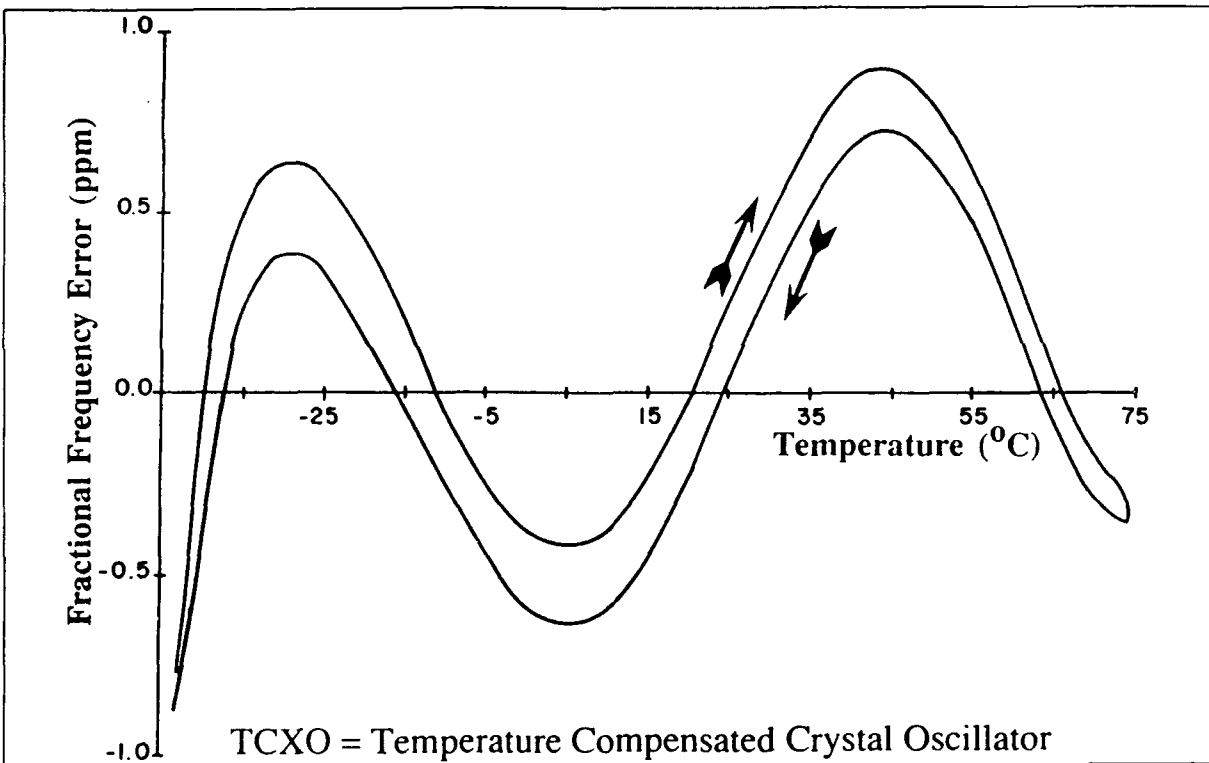


Figure 23. Temperature-compensated crystal oscillator (TCXO) thermal hysteresis, showing that the first f vs. T characteristic upon increasing temperature differs from the characteristic upon decreasing temperature.

Hysteresis is the major factor limiting the stability achievable with TCXOs. It is especially so in the MCXO because, in principle, the digital compensation method used in the MCXO would be capable of compensating for the f vs. T variations to arbitrary accuracy if the f vs. T characteristics could be described by single-valued functions. Retrace limits the accuracies achievable with OCXOs in applications where the OCXO is on-off cycled. Typical values of hysteresis in TCXOs range from 1 ppm to 0.1 ppm

when the temperature-cycling ranges are 0°C to 60°C, and -55°C to +85°C. Hysteresis of less than 1×10^{-8} has been observed in a few SC-cut (MCXO) resonators [24]. The typical MCXO resonator hysteresis in early models of the MCXO was a few parts in 10^8 [13]. Typical OCXO retrace specifications, after a 24 hour off period at about 25°C, range from 2×10^{-8} to 1×10^{-9} . Low-temperature storage during the off period, and extending the off period, usually make the retrace worse [16].

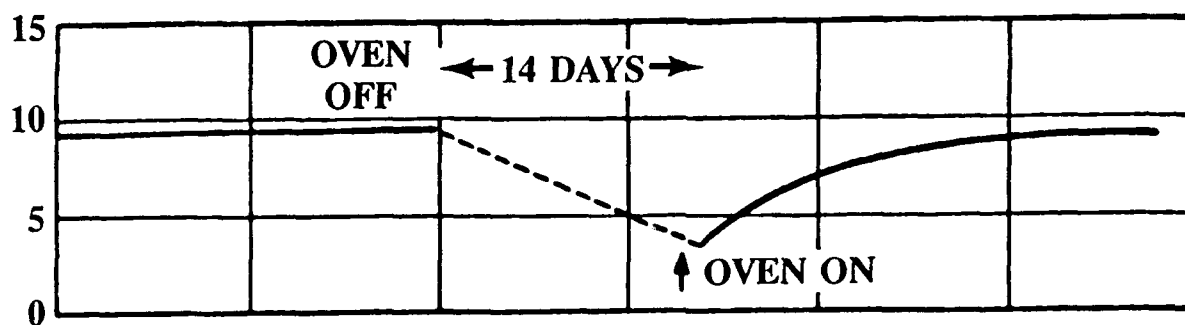


Figure 24. Oven-controlled crystal oscillator (OCXO) retrace example, showing that upon restarting the oscillator after a 14 day off-period, the frequency was about 7×10^{-9} lower than what it was just before turn-off, and that the aging rate had increased significantly upon the restart. About a month elapsed before the pre-turn-off aging rate was reached again.

The causes of hysteresis and retrace are not well understood; the experimental evidence to date is inconclusive [26]. The mechanisms that can cause these effects include strain changes, changes in the quartz, oscillator circuitry changes, contamination redistribution in the crystal enclosure, and apparent hysteresis or retrace due to thermal gradients.

E. Warm-up

When power is applied to a frequency standard, it takes a finite amount of time before the equilibrium frequency stability is reached. Figure 22, discussed above, illustrates the warm-up of two OCXOs. The warmup time of an oscillator is a function of the thermal properties of the resonator, the oscillator circuit and oven construction, the input power, and the oscillator's temperature prior to turn-on. Typical warm-up time specifications of OCXOs (e.g., from a 0°C start) range from 3 minutes to 10 minutes. Even TCXOs, MCXOs, and simple XOs take a few seconds to "warm up," although these are not ovenized. The reasons for the finite warm-up periods are that it takes a finite amount of time for the signal to build up in any high-Q circuit, and the few tens of milliwatts of power which are dissipated in these oscillators can change the thermal conditions within the oscillators.

F. Acceleration Effects

Acceleration changes a crystal oscillator's frequency [28]. The acceleration can be a steady-state acceleration, vibration, shock, attitude change (2-g tipover), or acoustic noise. The amount of frequency change depends on the magnitude and direction of the acceleration \vec{A} , and on the acceleration sensitivity of the oscillator $\vec{\Gamma}$. The acceleration sensitivity $\vec{\Gamma}$ is a vector quantity. The frequency change can be expressed as

$$\frac{\Delta f}{f} = \vec{\Gamma} \cdot \vec{A} .$$

Typical values of $|\vec{\Gamma}|$ are in the range of $10^{-9}/g$ to $10^{-10}/g$. For example, when $\vec{\Gamma} = 2 \times 10^{-9}/g$ and is normal to the earth's surface, and the oscillator is turned upside down (a change of 2 g), the frequency changes by 4×10^{-9} . When this oscillator is vibrated in the up-and-down direction, the time dependent acceleration modulates the oscillator's output frequency at the vibration frequency, with an amplitude of $2 \times 10^{-9}/g$.

When an oscillator is rotated 180° about a horizontal axis, the scalar product of the gravitational field and the unit vector normal to the initial "top" of the oscillator changes from -1g to +1g, i.e., by 2g. Figure 25 shows actual data of the fractional frequency shifts of an oscillator when the oscillator was rotated about three mutually perpendicular axes in the earth's gravitational field. For each curve, the axis of rotation was horizontal. The sinusoidal shape of each curve is a consequence of the scalar product being proportional to the cosine of the angle between the acceleration-sensitivity vector and the acceleration due to gravity [28].

In the frequency domain, the modulation results in vibration-induced sidebands that appear at plus and minus integer multiples of the vibration frequency from the carrier frequency. Figure 26 shows the output of a spectrum analyzer for a 10-MHz, $1.4 \times 10^{-9}/g$ oscillator that was vibrated at 100 Hz and 10 g. For sinusoidal vibration, the "sidebands" are spectral lines. When the frequency is multiplied, as it is in many applications, the sideband levels increase by 20 dB for each 10X multiplication. The increased sideband power is extracted from the carrier. Under certain conditions of multiplication, the carrier disappears, i.e., all the energy is then in the sidebands.

The acceleration sensitivity can be calculated from the vibration induced sidebands. The preferred method is to measure the sensitivity at a number of vibration frequencies in order to reveal resonances. Figure 27 shows an example of the results of a resonance in an OCXO. The resonance at 424 Hz amplified the acceleration sensitivity 17-fold.

The effect of random vibration is to raise the phase-noise level of the oscillator. The degradation of phase-noise can be substantial when the oscillator is on a vibrating platform, such as on an aircraft. Figure 28 shows a typical aircraft random-vibration

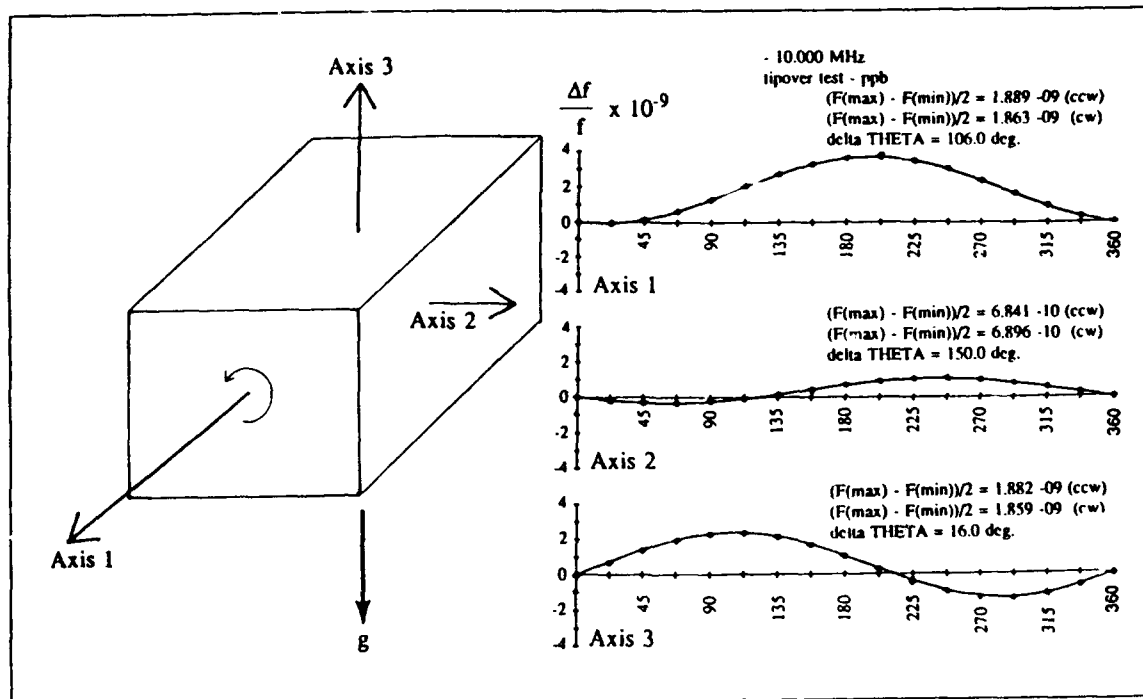


Figure 25. 2-g tipover test (Δf vs. attitude about three axes).

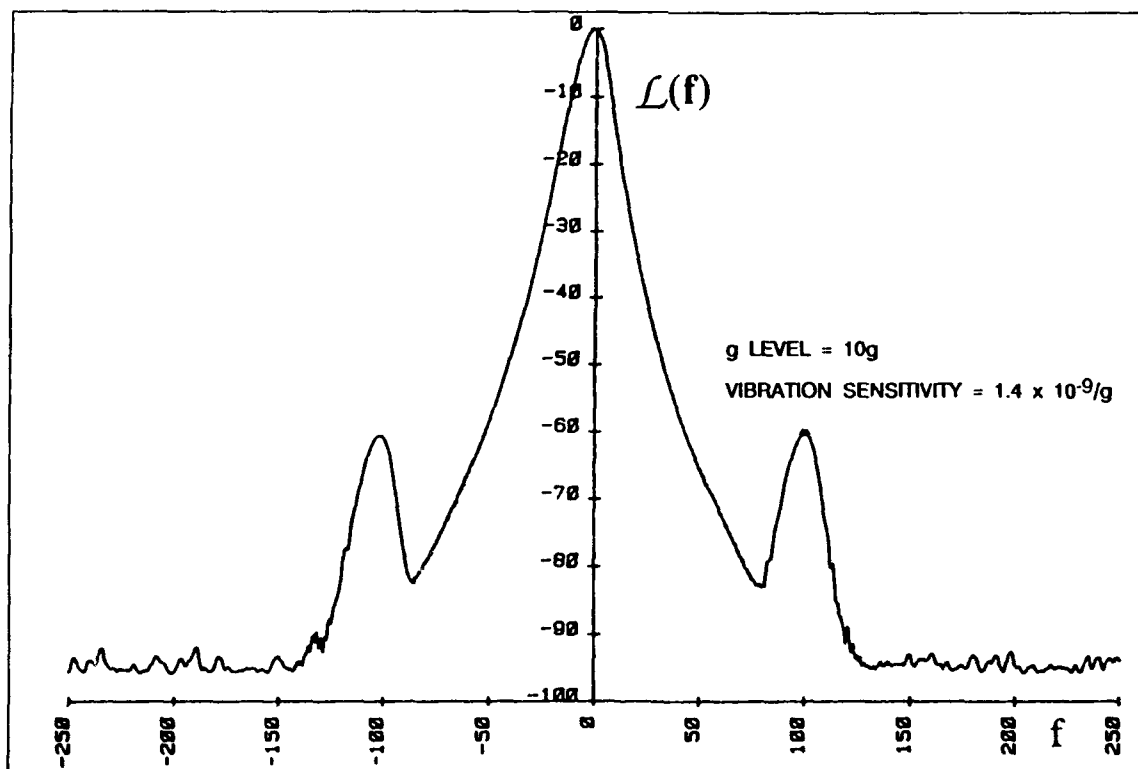


Figure 26. Vibration-induced "sidebands" (i.e., spectral lines).

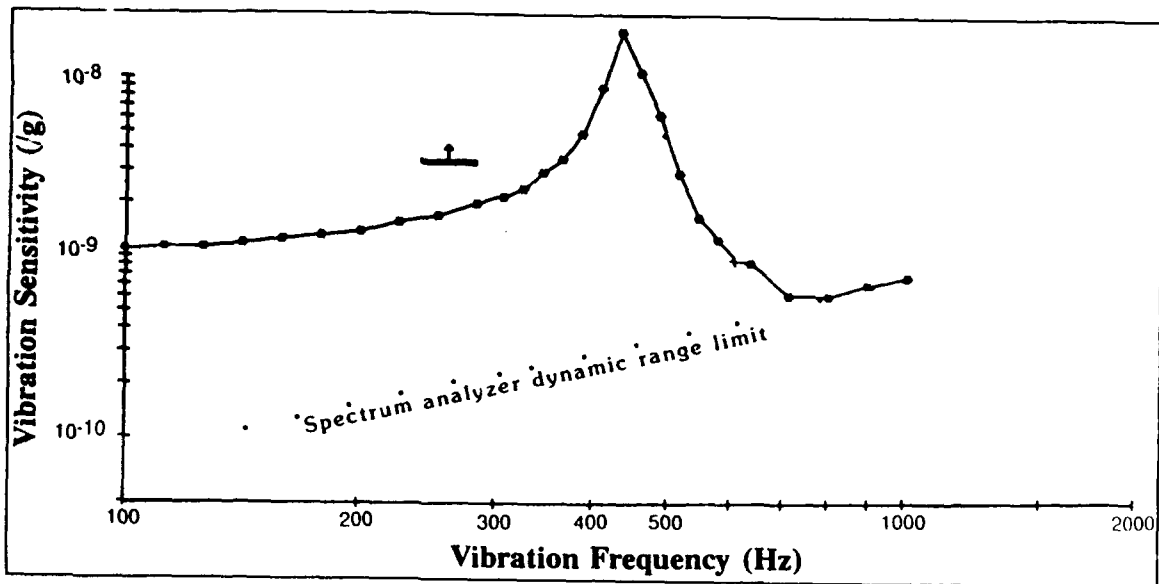


Figure 27. Resonance in the acceleration sensitivity vs. vibration frequency characteristic.

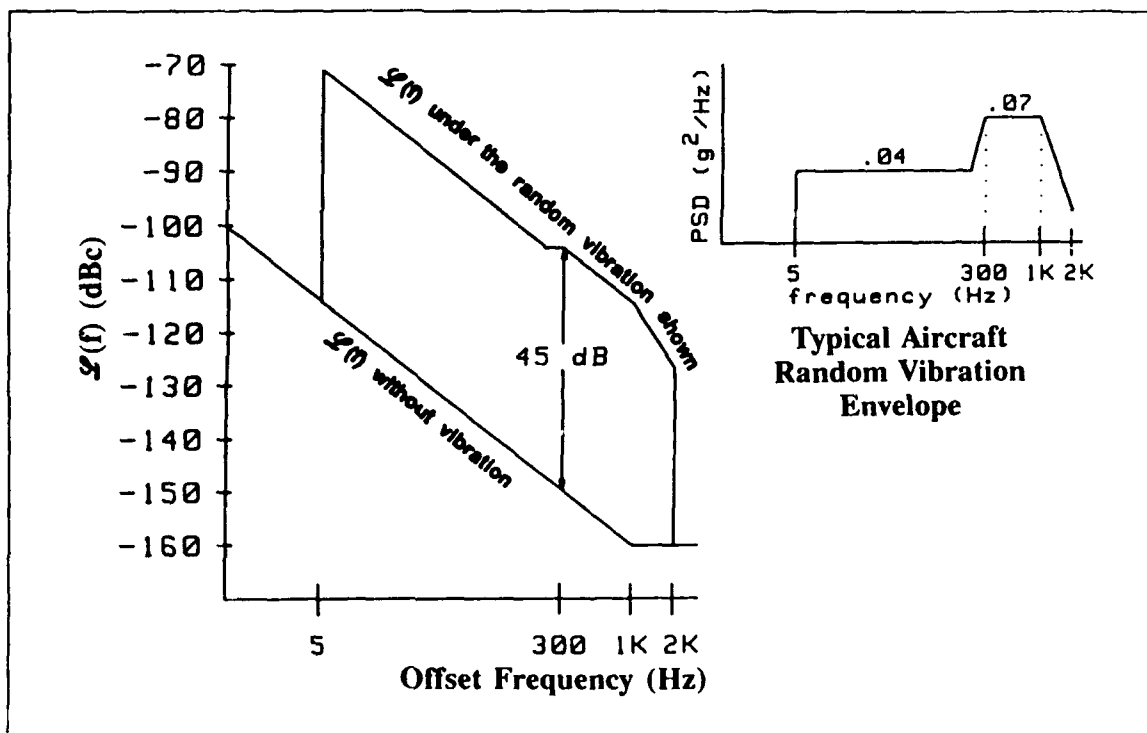


Figure 28. Random vibration induced phase noise degradation.

specification (power spectral density [PSD] vs. vibration frequency) and the resulting vibration-induced phase-noise degradation. Acoustic noise is another source of acceleration that can affect the frequency stability of oscillators.

The peak phase excursion, ϕ_{peak} , due to sinusoidal vibration is

$$\phi_{\text{peak}} = \frac{\Delta f}{f_v} = \frac{\vec{\Gamma} \cdot \vec{A}}{f_v} f_0 \text{ radians.}$$

To some system designers, the quantity of concern is the integrated phase noise in the band f_1 to f_2 , which is

$$\Phi^2 = 2 \int_{f_1}^{f_2} \mathcal{L}(f) df \text{ radians}^2.$$

Upon frequency multiplication, both ϕ_{peak} and Φ increase by the multiplication factor. For example, if $\vec{\Gamma} \cdot \vec{A} = 1 \times 10^{-9}$, $f_0 = 10$ MHz and $f_v = 10$ Hz, then $\phi_{\text{peak}} = 1 \times 10^{-3}$ radian. If this oscillator's frequency is multiplied to 10 GHz, e.g., in a radar system, then at 10 GHz, $\phi_{\text{peak}} = 1$ radian. Such large phase excursions can be catastrophic to many systems.

Figure 29 shows how the probability of detection for a coherent radar system varies with the phase noise of the reference oscillator [29]. The phase noise requirement for a 90% probability of detection of a 4km/hr target is -130 dBc per Hz at 70 Hz from the carrier, for a 10 MHz oscillator. Such a phase noise is well within the capability of 10 MHz oscillators, provided that the oscillators are in a quiet environment. However, when the oscillators are on a vibrating platform, such as an airborne radar system, the phase noise of even the best available oscillators (as of 1991) is degraded by an amount that reduces the probability of detection to zero.

During shock, a crystal oscillator's frequency changes suddenly due to the sudden acceleration, as is illustrated in Figure 30. The frequency change follows the expression above for acceleration-induced frequency change except, if during the shock some elastic limits in the crystal's support structure or electrodes are exceeded (as is almost always the case during typical shock tests), the shock will produce a permanent frequency change.

Permanent frequency offsets due to shock can also be caused by changes in the oscillator circuitry (e.g., due to movement of a wire or circuit board), and the removal of (particulate) contamination from the resonator surfaces. Resonances in the mounting structure will amplify the shock-induced stress.

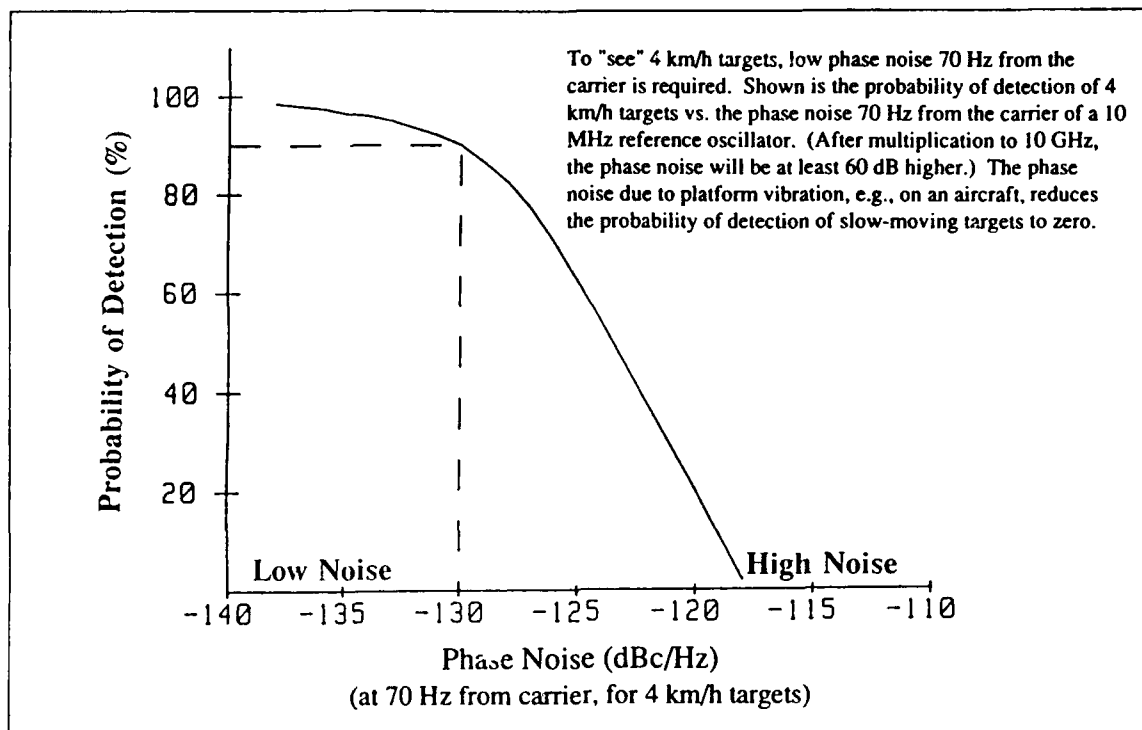


Figure 29. Coherent radar probability of detection as a function of reference oscillator phase noise.

If the shock level is sufficiently high, the crystal will break; however, in applications where high shock levels are a possibility, crystal units with chemically polished crystal plates can be used. Such crystals can survive shocks in excess of 30,000 g and have been fired successfully from howitzers [30,31].

G. Magnetic-Field Effects

Quartz is diamagnetic; however, magnetic fields can affect magnetic materials in the crystal unit's mounting structure, electrodes, and enclosure. Time-varying electric fields will induce eddy currents in the metallic parts. Magnetic fields can also affect components such as inductors in the oscillator circuitry. When a crystal oscillator is designed to minimize the effects of magnetic fields, the sensitivity can be much less than 10^{-10} per oersted. Magnetic-field sensitivities on the order of 10^{-12} per oersted have been measured in crystal units designed specifically for low magnetic-field sensitivity [32].

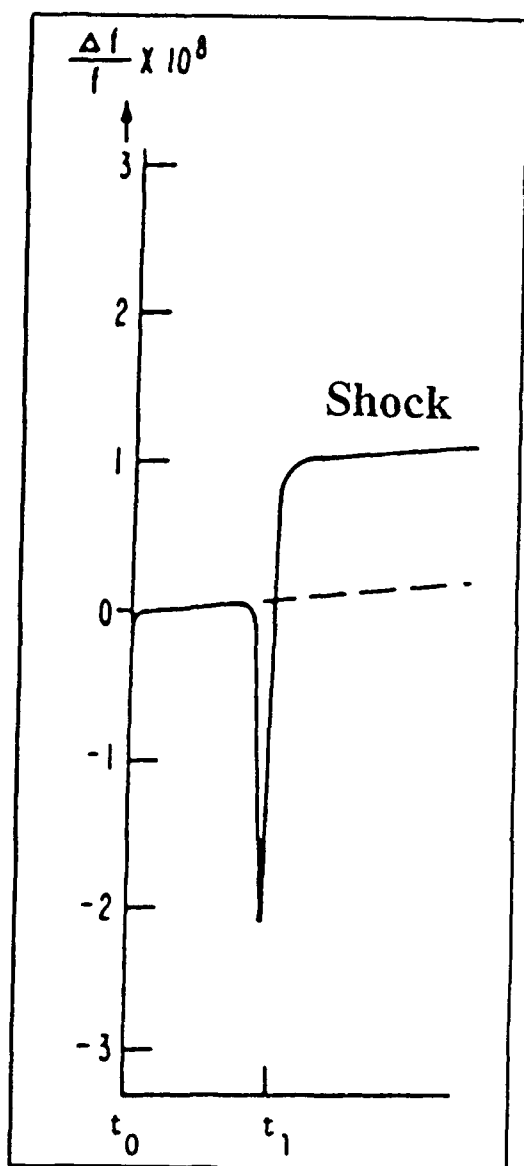


Figure 30. The effect of a shock at $t = t_1$ on oscillator frequency.

H. Radiation Effects

Ionizing radiation changes a crystal oscillator's frequency primarily because of changes the radiation produces in the crystal unit [33]. Under certain conditions, the radiation will also produce an increase in the crystal unit's equivalent series resistance. The resistance increase can be large enough to stop the oscillation when the oscillator is not radiation hardened.

Figure 31 shows a crystal oscillator's idealized frequency response to a pulse of ionizing radiation. The response consists of two parts. Initially, there is a transient frequency change that is due primarily to the thermal-transient effect caused by the sudden deposition of energy into the crystal unit. This effect is a manifestation of the dynamic f vs. T effect discussed above. The transient effect is absent in SC-cut resonators made of high purity quartz.

In the second part of the response, after steady state is reached, there is a permanent frequency offset that is a function of the radiation dose and the nature of the crystal unit. The frequency change versus dose is nonlinear, the change per rad being much larger at low doses than at large doses. At doses above 1 kilorad (Krad) (SiO_2), the rate of frequency change with dose is quartz impurity-defect dependent. For example, at a 1 megarad (Mrad) dose, the frequency change can be as large as 10 ppm when the crystal unit is made from natural quartz;

it is typically 1 to a few ppm when the crystal is made from cultured quartz, and it can be as small as 0.02 ppm when the crystal is made from swept cultured quartz.

The impurity defect of major concern in quartz is the substitutional Al^{3+} defect with its associated interstitial charge compensator, which can be an H^+ , Li^+ , or Na^+ ion, or a hole. This defect substitutes for a Si^{4+} in the quartz lattice. Radiation can result in a change in the position of weakly bound compensators, which changes the elastic constants of quartz and thereby leads to a frequency change. The movement of ions

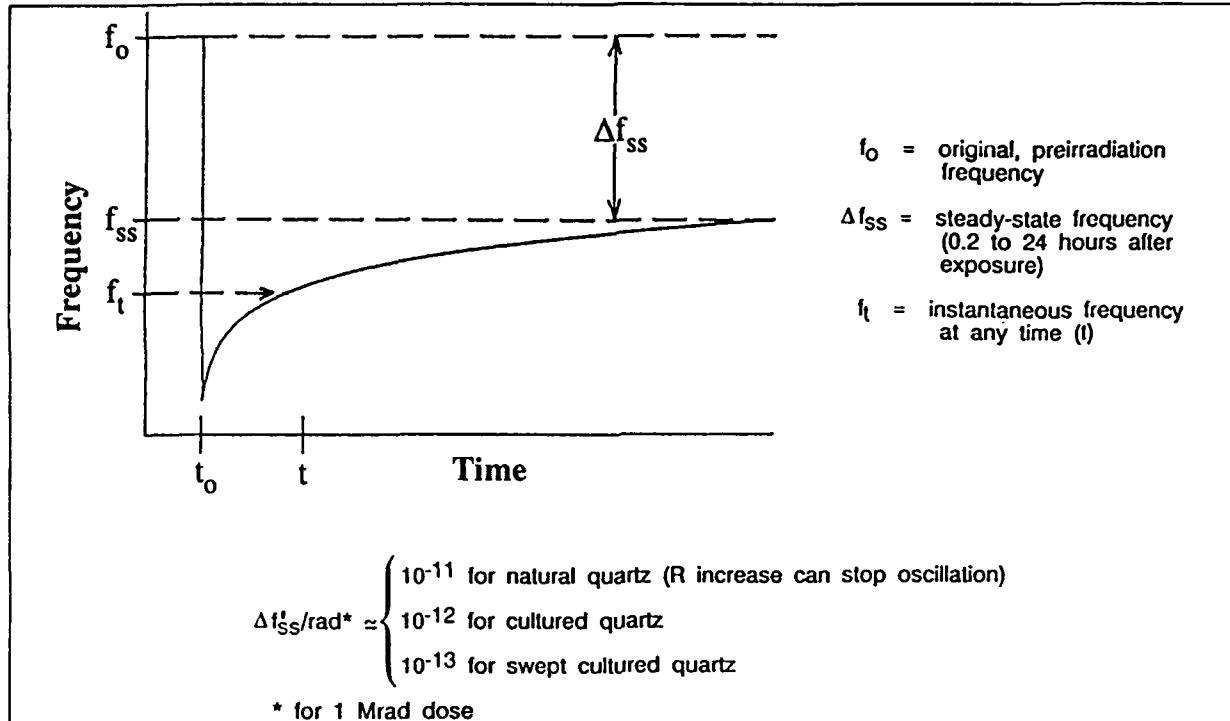


Figure 31. Crystal oscillator's response to a pulse of ionizing radiation: f_o = original, preirradiation frequency; Δf_{ss} = steady-state frequency offset (0.2 hours to 24 hours after exposure); f_t = instantaneous frequency at time t .

also results in a decrease in the crystal's Q , i.e., in an increase in the crystal's equivalent series resistance, especially upon exposure to a pulse of ionizing radiation. If the oscillator's gain margin is insufficient, the increased resistance can stop the oscillation for periods lasting many seconds. A high level pulse of ionizing radiation will produce photocurrents in the circuit which result in a momentary cessation of oscillation, independent of the type of quartz used in the resonator. In oscillators using properly designed oscillator circuitry and resonators made of swept quartz, the oscillator recovers within 15 μs after exposure [34,35].

Sweeping is a high-temperature, electric-field-driven, solid-state purification process in which the weakly bound alkali compensators are diffused out of the lattice and replaced by more tightly bound H^+ ions and holes [36,37]. In the typical sweeping process, conductive electrodes are applied to the Z surfaces of a quartz bar, the bar is heated to about 500°C , and a voltage is applied so as to produce an electric field of about 1 kilovolt per centimeter along the Z direction. After the current through the bar decays (due to the diffusion of impurities) to some constant value, the bar is cooled slowly, the voltage is removed, and then the electrodes are removed. Crystal units made from swept quartz

exhibit neither the radiation-induced Q degradation nor the large radiation-induced frequency shifts. Swept quartz (or low aluminum content quartz) should be used in oscillators which are expected to be exposed to ionizing radiation.

At low doses (e.g., at a few rads) the frequency change per rad can be as high as 10^{-9} per rad [38]. The low-dose effect is not well understood. It is not impurity-dependent, and it saturates at about 300 rads. At very high doses (i.e., at $\gg 1$ Mrad), the impurity-dependent frequency shifts also saturate because, since the number of defects in the crystal are finite, the effects of the radiation interacting with the defects are also finite.

When a fast neutron hurtles into a crystal lattice and collides with an atom, it is scattered like a billiard ball. A single such neutron can produce numerous vacancies, interstitials, and broken interatomic bonds. The effect of this "displacement damage" on oscillator frequency is dependent primarily upon the neutron fluence. The frequency of oscillation increases nearly linearly with neutron fluence at rates of: 8×10^{-21} neutrons per square centimeter (n/cm^2) at a fluence range of $10^{10} - 10^{12} n/cm^2$, $5 \times 10^{-21}/n/cm^2$ at $10^{12} - 10^{13} n/cm^2$, and $0.7 \times 10^{-21}/n/cm^2$ at $10^{17} - 10^{18} n/cm^2$.

I. Other Effects on Stability

Ambient pressure change (as during an altitude change) can change a crystal oscillator's frequency if the pressure change produces a deformation of the crystal unit's or the oscillator's enclosure (thus changing stray capacitances and stresses). The pressure change can also affect the frequency indirectly through a change in heat-transfer conditions inside the oscillator. Humidity changes can also affect the heat-transfer conditions. In addition, moisture in the atmosphere will condense on surfaces when the temperature falls below the dew point, and can permeate materials such as epoxies and polyimides, and thereby affect the properties (e.g., conductivities and dielectric constants) of the oscillator circuitry. The frequency of a properly designed crystal oscillator changes less than 5×10^{-9} when the environment changes from one atmosphere of air to a vacuum. The medium and long term stability of some oscillators can be improved by controlling the pressure and humidity around the oscillators [39].

Electric fields can change the frequency of a crystal unit. An ideal AT-cut is not affected by a DC voltage on the crystal electrodes, but "doubly rotated cuts," such as the SC-cut, are affected. For example, the frequency of a 5-MHz fundamental mode SC-cut crystal changes 7×10^{-9} per volt. Direct-current voltages on the electrodes can also cause sweeping, which can affect the frequencies of all cuts.

Power-supply and load-impedance changes affect the oscillator circuitry and, indirectly, the crystal's drive level and load reactance. A change in load impedance changes the amplitude or phase of the signal reflected into the oscillator loop, which

changes the phase (and frequency) of the oscillation [40]. The effects can be minimized through voltage regulation and the use of buffer amplifiers. The frequency of a "good" crystal oscillator changes less than 5×10^{-10} for a 10% change in load impedance. The typical sensitivity of a high-quality crystal oscillator to power-supply voltage changes is $5 \times 10^{-11}/V$.

Gas permeation under conditions where there is an abnormally high concentration of hydrogen or helium in the atmosphere can lead to anomalous aging rates. For example, hydrogen can permeate into "hermetically" sealed crystal units in metal enclosures, and helium can permeate through the walls of glass-enclosed crystal units.

J. Interactions Among the Influences on Stability

The various influences on frequency stability can interact in ways that lead to erroneous test results if the interfering influence is not recognized during testing. For example, building vibrations can interfere with the measurement of short-term stability. Vibration levels of 10^{-3} g to 10^{-2} g are commonly present in buildings. Therefore, if an oscillator's acceleration sensitivity is $1 \times 10^{-9}/g$, then the building vibrations alone can contribute short-term instabilities at the 10^{-12} to 10^{-11} level.

The 2-g tipover test is often used to measure the acceleration sensitivity of crystal oscillators. Thermal effects can interfere with this test because, when an oscillator is turned upside down, the thermal gradients inside the oven can vary due to changes in convection currents. Other examples of interfering influences include temperature and drive-level changes interfering with aging tests; induced voltages due to magnetic fields interfering with vibration-sensitivity tests; and the thermal-transient effect, humidity changes, and the effect of load-reactance temperature coefficient interfering with the measurement of crystal units' static f vs. T characteristics.

An important effect in TCXOs is the interaction between the frequency adjustment during calibration and the f vs. T stability [41]. This phenomenon is called the *trim effect*. In TCXOs, a temperature-dependent signal from a thermistor is used to generate a correction voltage that is applied to a varactor in the crystal network. The resulting reactance variations compensate for the crystal's f vs. T variations. During calibration, the crystal's load reactance is varied to compensate for the TCXO's aging. Since the frequency versus reactance relationship is nonlinear, the capacitance change during calibration moves the operating point on the frequency versus reactance curve to a point where the slope of the curve is different, which changes the compensation (i.e., compensating for aging degrades the f vs. T stability). Figure 32 shows how, for the same compensating C_L vs. T , the compensating f vs. T changes when the operating point is moved to a different C_L . Figure 33 shows test results for a 0.5 ppm TCXO that had a ± 6 ppm frequency-adjustment range (to allow for aging compensation for the life of the device). When delivered, this TCXO met its 0.5 ppm f vs. T specification; however, when

the frequency was adjusted ± 6 ppm during testing, the f vs. T performance degraded significantly.

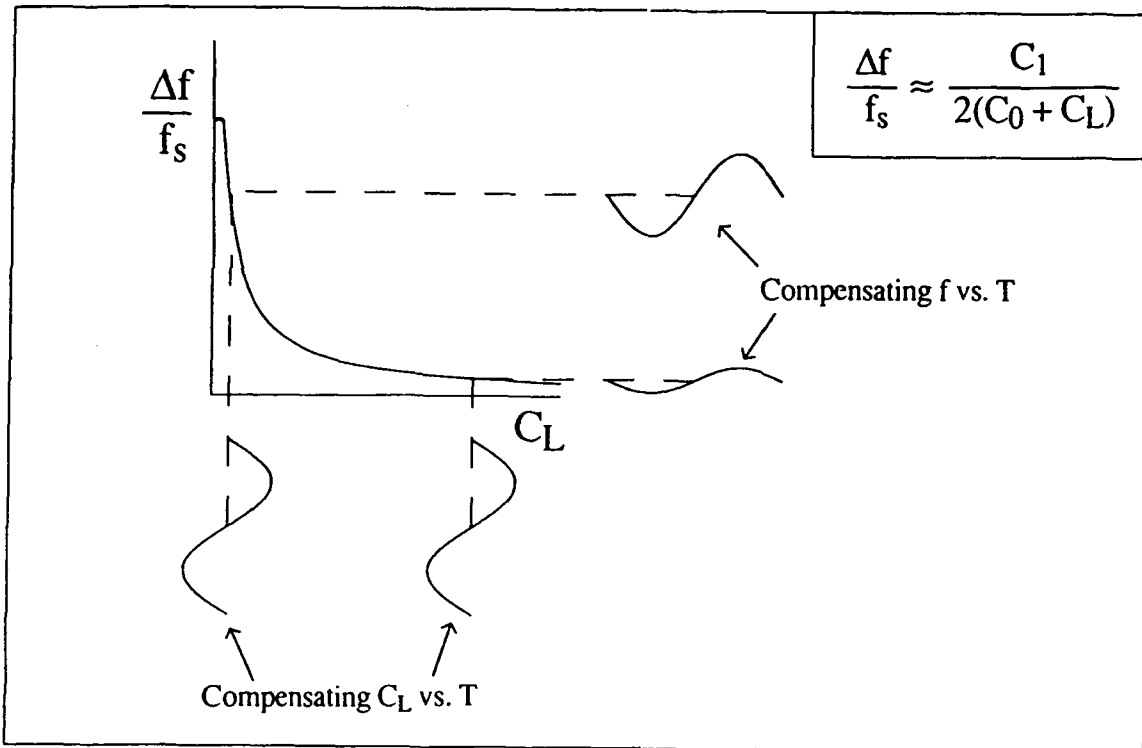


Figure 32. Change in compensating frequency versus temperature due to C_L change.

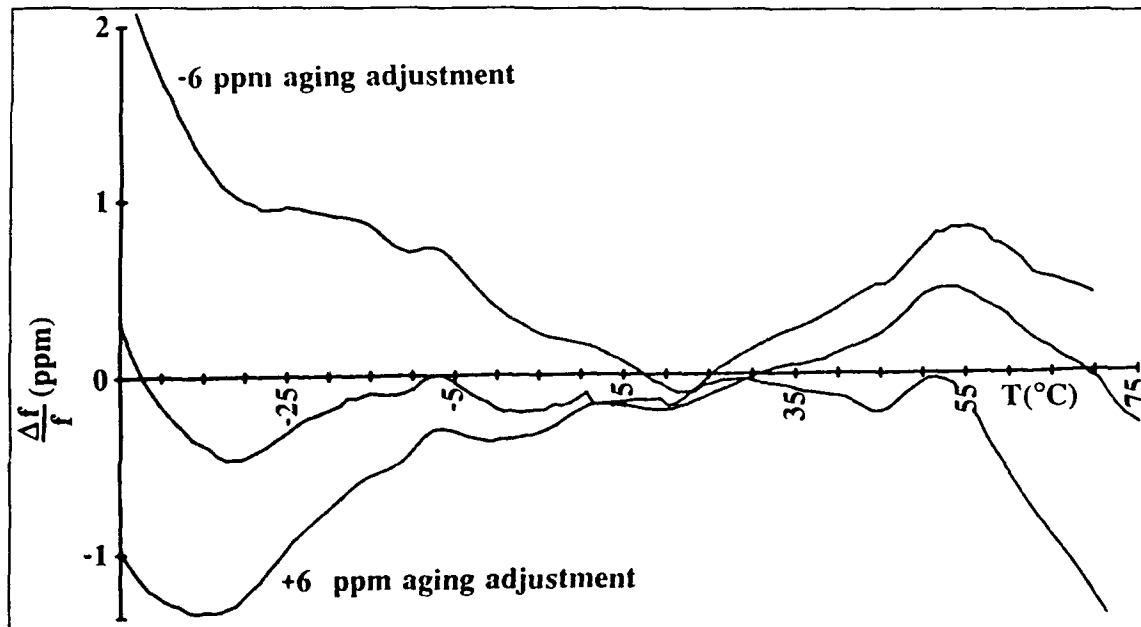
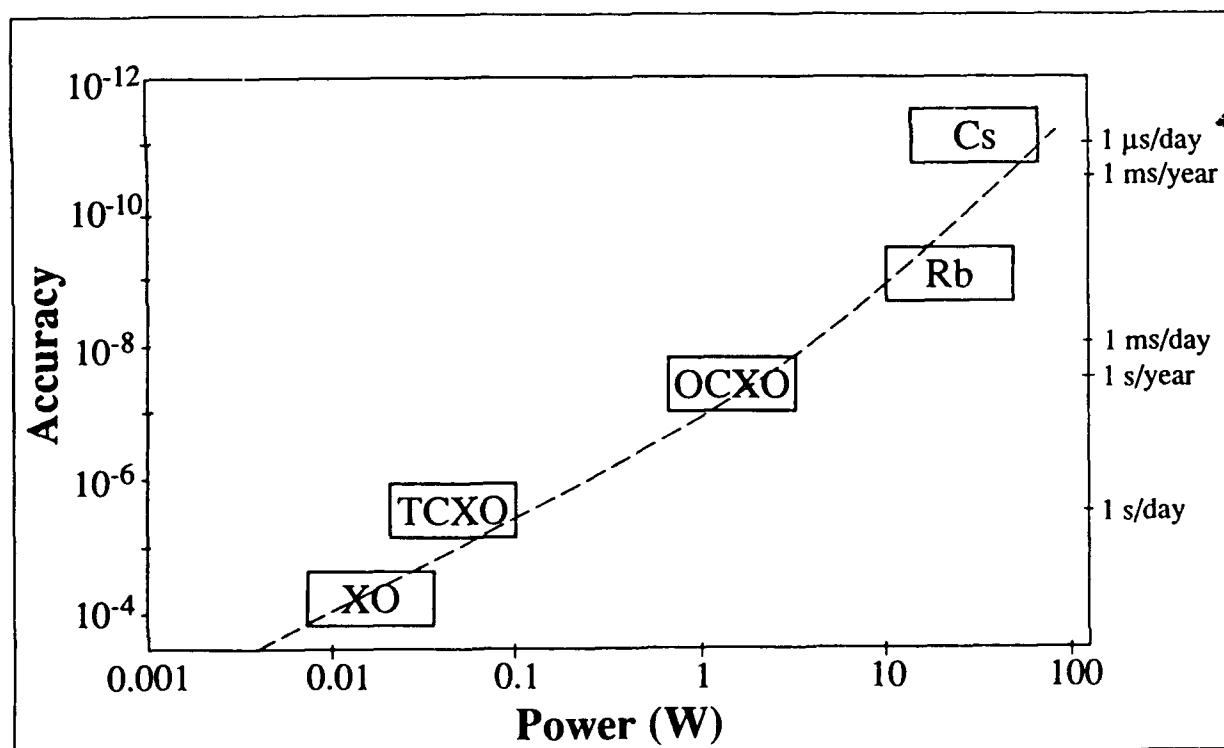


Figure 33. Temperature-compensated crystal oscillator (TCXO) trim effect.

IV. Oscillator Comparison and Selection

The discussion that follows applies to wide-temperature-range frequency standards (i.e., to those which are designed to operate over a temperature range that spans at least 90°C). Laboratory devices that operate over a much narrower temperature range can have better stabilities than those in the comparison below.

Commercially available frequency sources cover an accuracy range of several orders of magnitude--from the simple XO to the cesium-beam frequency standard. As the accuracy increases, so does the power requirement, size, and cost. Figure 34, for example, shows the relationship between accuracy and power requirement. Accuracy versus cost would be a similar relationship, ranging from about \$1 for a simple XO to about \$40,000 for a cesium standard (1991 prices). Table 1 shows a comparison of salient characteristics of frequency standards. Figure 35 shows the comparison of short term frequency stability ranges as a function of averaging time [42]. Figure 36 shows a comparison of phase-noise characteristics, and Table 2 shows a comparison of weaknesses and wear-out mechanisms.



* Accuracy vs. size, and accuracy vs. cost have similar relationships.

Figure 34. Relationship between accuracy and power requirements (XO=simple crystal oscillator; TCXO=temperature-compensated crystal oscillator; OCXO=oven-controlled crystal oscillator; Rb=rubidium frequency standard; Cs=cesium beam frequency standard).

Table 1. Salient characteristics comparison of frequency standards.

| | Quartz Oscillators | | | Atomic Oscillators | | |
|--|------------------------------------|------------------------------------|------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| | TCXO | MCXO | OCXO | Rubidium | RbXO | Cesium |
| Accuracy * (per year) | 2×10^{-6} | 5×10^{-8} | 1×10^{-8} | 5×10^{-10} | 7×10^{-10} | 2×10^{-11} |
| Aging/Year | 5×10^{-7} | 2×10^{-8} | 5×10^{-9} | 2×10^{-10} | 2×10^{-10} | 0 |
| Temp. Stab. (range, °C) | 5×10^{-7} (-55 to +85) | 2×10^{-8} (-55 to +85) | 1×10^{-9} (-55 to +85) | 3×10^{-10} (-55 to +68) | 5×10^{-10} (-55 to +85) | 2×10^{-11} (-28 to +65) |
| Stability, $\sigma_y(\tau)$ ($\tau = 1$ s) | 1×10^{-9} | 1×10^{-10} | 1×10^{-12} | 3×10^{-12} | 5×10^{-12} | 5×10^{-11} |
| Size (cm ³) | 10 | 50 | 20-200 | 800 | 1200 | 6000 |
| Warmup Time (min) | 0.1 (to 1×10^{-4}) | 0.1 (to 2×10^{-4}) | 4 (to 1×10^{-4}) | 3 (to 5×10^{-10}) | 3 (to 5×10^{-10}) | 20 (to 2×10^{-11}) |
| Power (W) (at lowest temp.) | 0.05 | 0.04 | 0.25 | 20 | 0.35 | 30 |
| Price (~\$) | 100 | 1,000 | 2,000 | 8,000 | 10,000 | 40,000 |

* Including environmental effects (note that the temperature ranges for Rb and Cs are narrower than for quartz).

Characteristics are provided in Table 1 for atomic oscillators: rubidium and cesium frequency standards and the rubidium-crystal oscillator (RbXO). In atomic frequency standards, the output signal frequency is determined by the energy difference between two atomic states, rather than by some property of a bulk material (as it is in quartz oscillators.) An introductory review of atomic frequency standards can be found in reference 43, and reference 44 is a review of the literature up to 1983. (Reference 43 reviews both atomic and quartz frequency standards; the report you are reading is based on the quartz portion of that reference.) The RbXO is a device intended for applications where power availability is limited, but where atomic frequency standard accuracy is needed [45]. It consists of a rubidium frequency standard, a low-power and high-stability crystal oscillator, and control circuitry that adjusts the crystal oscillator's frequency to that of the rubidium standard. The rubidium standard is turned on periodically (e.g., once a week) for the few minutes it takes for it to warm up and correct the frequency of the crystal oscillator. With the RbXO, one can approach the long-term stability of the rubidium standard with the low (average) power requirement of the crystal oscillator.

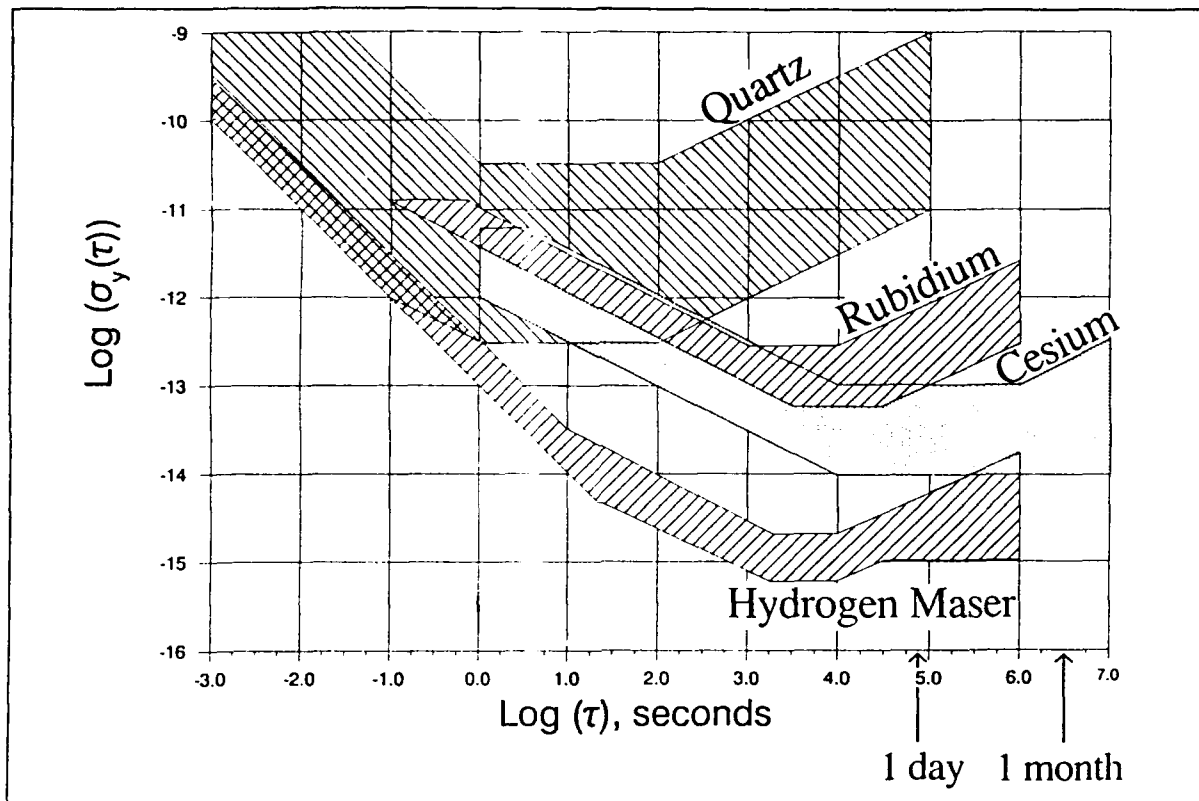
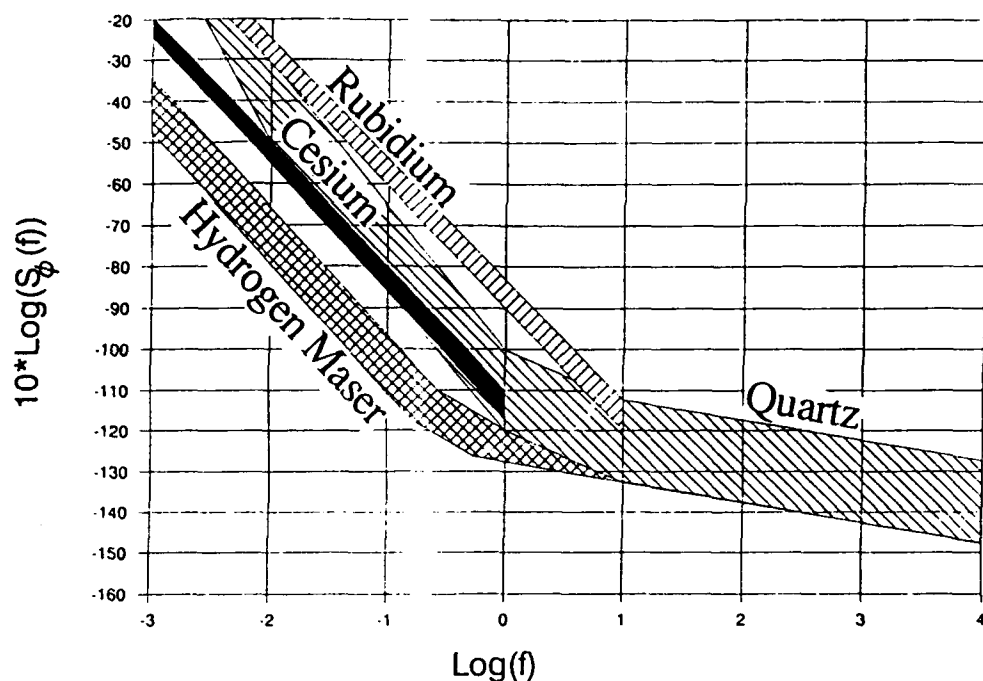


Figure 35. Stability as a function of averaging time comparison of frequency standards.

The major questions to be answered in choosing an oscillator include:

1. What frequency accuracy or reproducibility is needed for the system to operate properly?
2. How long must this accuracy be maintained, i.e., will the oscillator be calibrated or replaced periodically, or must the oscillator maintain the required accuracy for the life of the system?
3. Is ample power available, or must the oscillator operate from batteries?
4. What warmup time, if any, is permissible?
5. What are the environmental extremes in which the oscillator must operate?
6. What is the short-term stability (phase-noise) requirement?
7. What is the size constraint?

In relation to the second question, what cost is to be minimized: the initial acquisition cost or the life-cycle cost? Often, the cost of recalibration is far higher than the added cost of an oscillator that can provide calibration-free life. A better oscillator may also allow simplification of the system's design.



Typical One-sided Spectral Density of Phase Deviation vs. Offset Frequency, for Various Standards, Calculated at 5 MHz.

Figure 36. Phase instability comparison of frequency standards.

Table 2. Weaknesses and wear-out mechanisms comparison of frequency standards.

| | Weakness | Wearout Mechanisms |
|-----------------|--|---|
| Quartz | Aging Radiation hardness | None |
| Rubidium | Life Power Weight | Rubidium depletion Buffer gas depletion Glass contaminants |
| Cesium | Life Power Weight Cost Temperature range | Cesium supply depletion Spent cesium gettering Ion pump capacity Electron multiplier |

The frequency of the oscillator is another important consideration, because the choice can have a significant impact on both the cost and the performance. Everything else being equal, an oscillator of standard frequency, such as 5 MHz or 10 MHz, for which manufacturers have well established designs, will cost less than one of an unusual frequency, such as 8.34289 MHz. Moreover, for thickness-shear crystals, such as the AT cut and SC cut, the lower the frequency, the lower the aging [16]. Since at frequencies much below 5 MHz, thickness-shear crystals become too large for economical manufacturing, and since all the highest stability oscillators use thickness-shear crystals, the highest stability commercially available oscillator's frequency is 5 MHz. Such oscillators will also have the lowest phase-noise capability close to the carrier. There are also some excellent 10 MHz oscillators on the market; however, oscillators of much higher frequency than 10 MHz have significantly higher aging rates and phase-noise levels close to the carrier than do 5 MHz oscillators. For lowest phase-noise far from the carrier, where the signal-to-noise ratio determines the noise level, higher frequency crystals (e.g., 100 MHz) can provide lower noise because such crystals can tolerate higher drive levels, thereby allowing higher signal levels.

V. Specifications, Standards, Terms, and Definitions

Numerous specifications and standards exist which relate to frequency standards. The major organizations responsible for these documents are the Institute of Electrical and Electronics Engineers (IEEE), the International Electrotechnical Commission (IEC), the CCIR, and the U. S. Department of Defense, which maintains the Military Specification (MIL-SPEC) system. A listing of "Specifications and Standards Relating to Frequency Control" can be found in the final pages of the *Proceedings of the Annual Symposium on Frequency Control*. In the 1990 *Proceedings*, for example, 79 such documents are listed [46]. Many of the documents include terms and definitions, some of which are inconsistent. Unfortunately, no single authoritative document exists for terms and definitions relating to frequency standards. The terms and definitions in the CCIR glossary [15], in IEEE Std. 1139-1988 [18], and in MIL-O-55310's section 6 [27] are the most recent; they address different aspects of the field, and together form a fairly good set of terms and definitions for users of frequency standards.

The most comprehensive document dealing with the specification of frequency standards is MIL-O-55310 [27]. The evolution of this document over a period of many years has included periodic coordinations between the government agencies that purchase crystal oscillators and the suppliers of those oscillators. The document addresses the specifications of all the oscillator parameters discussed above, plus many others. This specification was written for crystal oscillators. Because the output frequencies of atomic frequency standards originate from crystal oscillators, and because no comparable document exists that addresses atomic standards specifically, MIL-O-55310 can also serve as a useful guide to specifying atomic standards.

MIL-STD-188-115, *Interoperability and Performance Standards for Communications Timing and Synchronization Subsystems*, specifies that the standard frequencies for nodal clocks shall be 1 MHz, 5 MHz, or 5×2^N MHz, where N is an integer. This standard also specifies a 1-pulse-per-second timing signal of amplitude 10 V, pulse width of 20 μ s, rise time less than 20 ns, fall time less than 1 μ s; and a 24-bit binary coded decimal (BCD) time-code that provides Coordinated Universal Time (UTC) time of day in hours, minutes, and seconds, with provisions for an additional 12 bits for day of the year, and an additional four bits for describing the figure of merit (FOM) of the time signal. The FOMs range from BCD Character 1 for better than 1 ns accuracy to BCD character 9 for "greater than 10 ms of fault" [47].

VI. For Further Reading

Reference 4 contains a thorough bibliography on the subject of frequency standards to 1983. The principal forum for reporting progress in the field has been the *Proceedings of the Annual Symposium on Frequency Control* [46]. Other publications that deal with frequency standards include *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, *IEEE Transactions on Instrumentation and Measurement*, *Proceedings of the Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting* [48], and *Proceedings of the European Frequency and Time Forum* [49]. Review articles can be found in special issues and publications [50-54].

VII. References

1. Smith, W. L., Precision Oscillators. In: *Precision Frequency Control*, Vol. 2 (E. A. Gerber and A. Ballato, eds.), Academic Press, New York, 1985, pp. 45-98.
2. Frerking, M. E., *Crystal Oscillator Design and Temperature Compensation*, Van Nostrand Reinhold Company, 1978.
3. Bottom, V. E., *Introduction to Quartz Crystal Unit Design*, Van Nostrand Reinhold, New York, 1982.
4. Gerber, E. A. and Ballato, A. (eds.), *Precision Frequency Control*, Academic Press, New York, 1985.
5. Parzen, B., *Design of Crystal and Other Harmonic Oscillators*, Wiley, New York, 1983. Chapter 3 of this book, Piezoelectric Resonators, by A. Ballato, is an oscillator-application oriented treatment of the subject.
6. Hafner, E., Resonator and Device Measurements. In: *Precision Frequency Control*,

Vol. 2 (E. A. Gerber and A. Ballato, eds.), Academic Press, 1985, pp. 1-44.

7. Benjaminson, A., Computer-Aided Design of Crystal Oscillators, U. S. Army Laboratory Command R&D Technical Report DELET-TR-84-0386-F, August 1985, AD-B096820; Advanced Crystal Oscillator Design, U. S. Army Laboratory Command R&D Technical Report SLCET-TR-85-0445-F, January 1988, AD-B121288; Advanced Crystal Oscillator Design, U. S. Army Laboratory Command R&D Technical Report SLCET-TR-88-0804-1, February 1989, AD-B134514.
8. Kusters, J. A., Resonator and Device Technology. In: *Precision Frequency Control*, Vol. 1 (E. A. Gerber and A. Ballato, eds.), Academic Press, 1985, pp. 161-183.
9. Meeker, T. R., Theory and Properties of Piezoelectric Resonators and Waves. In: *Precision Frequency Control*, Vol. 1 (E. A. Gerber and A. Ballato, eds.), Academic Press, 1985, pp. 47-119.
10. Kusters, J. A., The SC-cut Crystal - An Overview, *Proc. 1981 Ultrasonics Symp.* pp. 402-409.
11. R. L. Filler, The Amplitude-Frequency Effect in SC-cut Resonators, *Proc. 39th Ann. Symp. on Frequency Control*, pp. 311-316, 1985, IEEE Cat. No. 85CH2186-5.
12. Frerking, M. E., Temperature Control and Compensation. In: *Precision Frequency Control*, Vol. 2 (E. A. Gerber and A. Ballato, eds.), Academic Press, New York, 1985, pp. 99-111.
13. Schodowski, S. S., et al., Microcomputer Compensated Crystal Oscillator for Low Power Clocks, *Proc. 21st Ann. Precise Time & Time Interval (PTTI) Applications & Planning Meeting*, 445-464 (1989). Available from the U. S. Naval Observatory, Time Services Department, 34th and Massachusetts Avenue, NW, Washington, DC 20392. Details of the MCXO are also described in a series of five papers in the *Proc. 43rd Ann. Symp. Frequency Control*, IEEE Catalog No. 89CH2690-6 (1989).
14. XIIIth General Conference of Weights and Measures, Geneva, Switzerland, October 1967.
15. International Radio Consultative Committee (CCIR), Recommendation No. 686, Glossary. In: *CCIR 17th Plenary Assembly*, Vol. 7, Standard Frequencies and Time Signals (Study Group 7), CCIR, Geneva, Switzerland, 1990. Copies available from International Telecommunications Union, General Secretariat - Sales Section, Place des Nations, CH1211 Geneva, Switzerland.
16. Vig, J. R., and Meeker, T. R., The Aging of Bulk Acoustic Wave Resonators, Filters, and Oscillators, *Proc. 45th Ann. Symp. Frequency Control*, pp. 77-101, IEEE Cat. No. 91CH2965-2, (1991).

17. Rutman, J., and Walls, F. L., Characterization of Frequency Stability in Precision Frequency Sources, *Proc. of the IEEE*, Vol. 79, No. 6, pp. 952-960, June 1991.
18. IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology, *IEEE Std. 1139-1988*. IEEE, 445 Hoes Lane, Piscataway, NJ 08854, U.S.A.
19. Parker, T. E., Characteristics and Sources of Phase Noise in Stable Oscillators, *Proc. 41st Ann. Symp. on Frequency Control*, pp. 99-110, 1987, IEEE Cat. No. 87CH2427-3.
20. Ballato, A. and Lukaszek, T., Higher Order Temperature Coefficients of Frequency of Mass-Loaded Piezoelectric Crystal Plates, *Proc. 29th Ann. Symp. on Frequency Control*, pp. 10-25, 1975, AD-A017466.
21. Ballato, A. and Vig, J. R., Static and Dynamic Frequency-Temperature Behavior of Singly and Doubly Rotated, Oven-Controlled Quartz Resonators, *Proc. 32nd Ann. Symp. on Frequency Control*, pp. 180-188, 1978, AD-A955718.
22. Ballato, A. and Tilton, R., Electronic Activity Dip Measurement, *IEEE Trans. on Instrumentation and Measurement*, Vol. IM-27, pp. 59-65, 1978.
23. Ballato, A., Frequency-Temperature-Load Capacitance Behavior of Resonators for TCXO Applications, *IEEE Trans. Sonics and Ultrasonics*, SU-25, pp. 185-191, July 1978.
24. Filler, R. L., Frequency-Temperature Considerations for Digital Temperature Compensation, *Proc. 45th Ann. Symp. on Frequency Control*, pp. 398-404, 1991, IEEE Cat. No. 87CH2965-2.
25. Filler, R. L., Measurement and Analysis of Thermal Hysteresis in Resonators and TCXOs, *Proc. 42nd Ann. Symp. Frequency Control*, pp. 380-388, 1988.
26. Kusters, J. A., and Vig, J. R., Thermal Hysteresis in Quartz Resonators - A Review, *Proc. 44th Ann. Symp. Frequency Control*, 165-175, IEEE Catalog No. 90CH2818-3 (1990).
27. U.S. Department of Defense, Military Specification, Oscillators, Crystal, General Specification for, MIL-O-55310. The latest revision is available from Military Specifications and Standards, 700 Robbins Ave., Bldg. 4D, Philadelphia, PA 19111-5094.
28. Vig, J. R., et al., *The Effects of Acceleration on Precision Frequency Sources (Proposed for IEEE Standards Project P1193)*, U.S. Army Laboratory Command Research and Development Technical Report SLCET-TR-91-3, March 1991. Copies available from National Technical Information Service, 5285 Port Royal Road, Sills Building, Springfield, VA 22161; NTIS accession no. AD-A235470.

29. Taylor, J., Effects of Crystal Reference Oscillator Phase Noise in a Vibratory Environment, Technical Memorandum 2799-1011, Motorola, SOTAS Engineering Development Section, Radar Operations, August 8, 1980. The Stand-Off Target Acquisition System (SOTAS) was a helicopter-borne coherent radar system that was being developed for the U.S. Army; the program was canceled before completion. To meet the 90% probability of detection goal for 4 km/h targets while operating from a helicopter, the system required a reference oscillator acceleration sensitivity of 4×10^{-12} per g, which was about 100-fold beyond the state-of-the-art at the time.
30. Vig, J. R., LeBus, J. W., and Filler, R. L., Chemically Polished Quartz, *Proc. 31st Ann. Symp. on Frequency Control*, pp. 131-143, 1977, AD-A088221.
31. Filler, R. L., et al., Ceramic Flatpack Enclosed AT and SC-cut Resonators, *Proc. 1980 IEEE Ultrasonic Symp.*, 819-824 (1980).
32. Brendel, R., et al., Influence of Magnetic Field on Quartz Crystal Oscillators, *Proc. 43rd Ann. Symp. Frequency Control*, 268-274, IEEE Catalog No. 89CH2690-6 (1989).
33. King, J. C., and Koehler, D. R., Radiation Effects on Resonators. In: *Precision Frequency Control*, Vol. 2 (E. A. Gerber and A. Ballato, eds.), Academic Press, New York, 1985, pp. 147-159.
34. King, J. C. and Sander, H. H., Rapid Annealing of Frequency Change in High Frequency Crystal Resonators Following Pulsed X-irradiation at Room Temperature, *Proc. 27th Ann. Symp. Frequency Control*, 117-119, NTIS accession no. AD-771042 (1973).
35. Paradysz, R. E. and Smith W. L., Crystal Controlled Oscillators for Radiation Environments, *Proc. 27th Ann. Symp. Frequency Control*, 120-123, NTIS accession no. AD-771042 (1973).
36. Martin, J. J., Electrodiffusion (Sweeping) of Ions in Quartz, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 35, No. 3, pp. 228-296, May 1988, IEEE Catalog 88CH2588-2.
37. Gualtieri, J. G., Sweeping Quartz Crystals, *Proc. 1989 IEEE Ultrasonics Symp.*, pp. 381-391, 1989.
38. Flanagan, T. M., Leadon, R. E., and Shannon, D. L., Evaluation of Mechanisms for Low-Dose Frequency Shifts in Crystal Oscillators, *Proc. 40th Ann. Symp. Frequency Control*, 127-133, NTIS accession no. AD-A235435 (1986).
39. Walls, F. L., The Influence of Pressure and Humidity on the Medium and Long-Term Frequency Stability of Quartz Oscillators, *Proc. 42nd Ann. Symp. on Frequency Control*, pp. 279-283, 1988, IEEE Cat. No. 88CH2588-2.

40. Walls, F. L., Environmental Sensitivities of Quartz Crystal Oscillators, *Proc. 22nd Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 465-477, 1990, AD-A239372.
41. Filler, et al., Specification and Measurement of the Frequency Versus Temperature Characteristics of Crystal Oscillators, *Proc. 43rd Ann. Symp. Frequency Control*, 253-256, IEEE Catalog No. 89CH2690-6 (1989).
42. The graphs in Figures 34 and 35 were prepared in 1989 for a CCIR document by Richard Sydnor, Jet Propulsion Laboratories.
43. Stein, S. R. and Vig, J. R., Communications Frequency Standards, in: *The Froelich/Kent Encyclopedia of Telecommunications*, Vol. 3 (Froehlich, F. E. and Kent, A. eds.), Marcel Dekker, Inc., New York, 1992, pp. 445-500. A reprint of this chapter is available under the title "Frequency Standards for Communications," as U. S. Army Laboratory Command Technical Report SLCET-TR-91-2 (Rev. 1), October 1991, NTIS Accession No. AD-A243211.
44. Hellwig, H., Microwave Time and Frequency Standards, In: *Precision Frequency Control*, Vol. 2 (E. A. Gerber and A. Ballato, eds.), Academic Press, New York, 1985, pp. 113-176.
45. Riley, W. J. and Vaccaro, J. R., A Rubidium-Crystal Oscillator (RbXO), *IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control*, Vol. UFFC-34, pp. 612-618, 1987.
46. The Proceedings of the Annual Symposium on Frequency Control have been published since the tenth symposium in 1956. The earlier volumes are available from the National Technical Information Service, 5285 Port Royal Road, Sills Building, Springfield, VA 22161, USA; the later volumes, from the IEEE, 445 Hoes Lane, Piscataway, NJ 08854. Ordering information for all the *Proceedings* can be found in the back of the latest volumes (e.g., the *Proceedings of the 45th Annual Symposium on Frequency Control* 1991, is available from the IEEE, Cat. No. 91CH2965-2).
47. U.S. Department of Defense, Military Standard, MIL-STD-188-115, *Interoperability and Performance Standards for Communications Timing and Synchronization Subsystems*. The latest revision is available from Military Specifications and Standards, 700 Robbins Avenue, Building 4D, Philadelphia, PA 19111-5094.
48. *The Proceedings of the Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting* are available from the U. S. Naval Observatory, Time Services Department, 34th and Massachusetts Ave., N.W., Washington, DC 20392-5100. The latest volumes are also available from the National Technical Information Service, 5285 Port Royal Road, Sills Building, Springfield, VA 22161, USA.

49. The *Proceedings of the European Frequency and Time Forum* are available from the Swiss Foundation for Research in Metrology (FSRM), Rue de l'Orangerie 8, CH-2000 Neuchatel, Switzerland.
50. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control*, UFFC-34 (November 1987).
51. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control*, UFFC-35 (May 1988).
52. Kroupa, V. F. (ed.), *Frequency Stability: Fundamentals and Measurement*, IEEE Press, New York, 1983.
53. Sullivan, D. B., et al. (eds.), *Characterization of Clocks and Oscillators*, National Institute of Standards and Technology Technical Note 1337, National Institute of Standards and Technology, Boulder, CO 80303-3328.
54. *Proceedings of the IEEE*, Special Issue on Time and Frequency, Vol. 79, No. 7, July 1991.

TUTORIAL ON HIGH PERFORMANCE ANALOG FIBER OPTIC SYSTEMS*

JPL



GEORGE F. LUTES AND RONALD T. LOGAN

**JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA**

* This work represents the results of one phase of research carried out by the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration.

INTRODUCTION

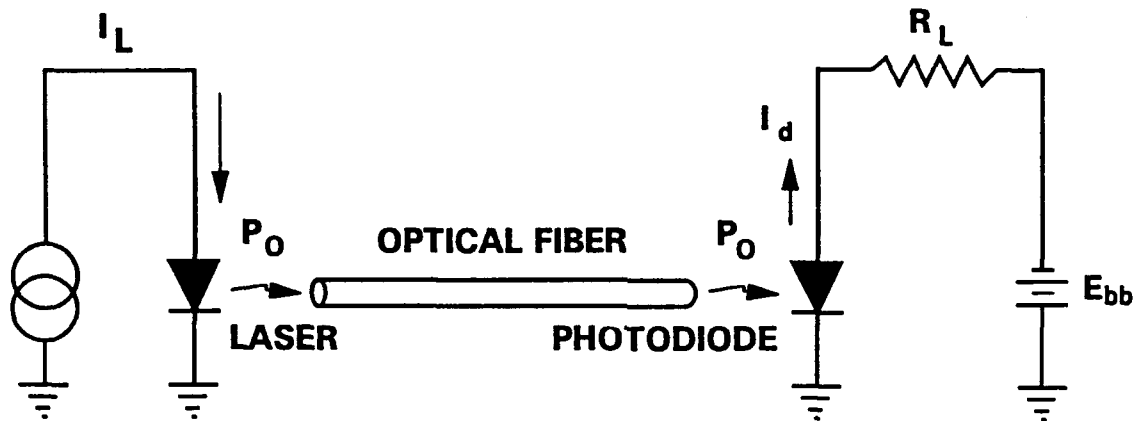
JPL



- **THIS TUTORIAL WILL:**
 - **DESCRIBE THE FUNDAMENTALS OF ANALOG FIBER OPTIC TRANSMISSION**
 - **DISCUSS NOISE SOURCES**
 - **DISCUSS CONDITIONS FOR BEST PERFORMANCE AND LIMITATIONS**
 - **GIVE THE PERFORMANCE OF TODAY'S ANALOG FIBER OPTIC SYSTEMS**
 - **PREDICT THE PERFORMANCE OF FUTURE ANALOG FIBER OPTIC SYSTEMS**

FIBER OPTIC TRANSMISSION SYSTEM MODEL

JPL



Fiber Optic Transmission System Model

The block diagram of a basic fiber optic system is given here. Electrical current (I_L) from a current source is applied to the laser diode. This current generates photons in the laser diode and the laser emits this optical power (P_O) into an optical fiber. The photons travel through the optical fiber and illuminate the photodiode at the far end.

The photodiode is biased through a load resistor (R_L) by a voltage source (E_{bb}). Current flows through the photodiode in proportion to the optical power incident on it.

FIBER OPTIC TRANSMISSION SYSTEM MODEL DEFINITIONS

I_L = the signal current applied to the laser diode

P_o = the optical power emitted by the laser diode

I_d = the photodetector current

R_L = the photodetector's load resistor

E_{bb} = the photodetector bias supply

P_e = the electrical power dissipated in the load resistor

η_L = the laser efficiency factor

η_d = the photodiode efficiency

λ = the wavelength of the laser

h = Planck's constant

c = the speed of light

Fiber Optic Transmission System Model Definitions

The definitions given here will be used in the following calculation of the relationship between optical signal power and the detected electrical signal power.

A further definition which was inadvertently left out is e = the electron charge.

FIBER OPTIC TRANSMISSION SYSTEM MODEL (CONT.)

$$P_o = \frac{I_L \eta_L h c}{e \lambda} \quad (1)$$

$$I_d = \frac{e \eta_d \lambda}{h c} P_o \text{ or } P_o = \frac{I_d h c}{e \eta_d \lambda} \quad (2)$$

From (1) and (2),

$$\frac{I_d h c}{e \eta_d \lambda} = \frac{I_L \eta_L h c}{e \lambda} \quad \therefore I_d = \eta_L \eta_d I_L \quad (3)$$

$$P_e = I_d^2 R_L \quad (4)$$

$$\text{From (2) and (4), } P_e = \left(\frac{e \eta_d \lambda}{h c} P_o \right)^2 R_L \quad (5)$$

$$\therefore P_e \propto P_o^2 \quad (6)$$

Fiber Optic Transmission System Model (cont.)

These calculations show that the detected electrical signal power is the square of the optical signal power. The importance of this will be shown shortly.

TERMINOLOGY

JPL



- dBe = electrical power ratio
- dBo = optical power ratio
- dBme = electrical power relative to 1 mW
- dBmo = optical power relative to 1 mW

Terminology

The terminology given here is used by the optical systems R&D and engineering communities to distinguish between optical power and electrical power and optical and electrical power ratios. The importance of this distinction is demonstrated in the next viewgraph.

IMPLICATIONS OF $P_e \propto P_o^2$

JPL



- ELECTRICAL SIGNAL-TO-NOISE RATIO (SNR) IS THE SQUARE OF THE OPTICAL SNR
- OPTICAL COUPLERS HAVING 52 dB_o ISOLATION HAVE 104 dB_e ISOLATION
- OPTICAL SWITCHES HAVING 60 dB_o ISOLATION HAVE 120 dB_e ISOLATION
- OPTICAL ISOLATORS HAVING 40 dB_o ISOLATION HAVE 80 dB_e ISOLATION

Implications of $P_e \propto P_o^2$

Because of the squared relationship between electrical signal power and optical signal power the electrical signal-to-noise-ratio is the square of the optical signal-to-noise-ratio, and the equivalent electrical signal isolation in an electro-optic system is the square of the optical isolation. For instance, optical couplers having a typical optical signal isolation of 52 dB_o have an equivalent electrical signal isolation of 104 dB_e. Optical switches having a typical optical signal isolation of > 60 dB_o have an equivalent electrical isolation of > 120 dB_e. Finally, optical isolators having a typical isolation of 40 dB_o have an equivalent electrical signal isolation of 80 dB_e.

OPTICAL SINUSOIDAL INTENSITY MODULATION

JPL



$$P(t) = P_o (1 + m \cos \omega t) + S(t)$$

where,

P_o = the average optical power (W),

m = the modulation index,

ω = the angular modulation frequency (rad/s), and

$S(t)$ = is the noise generated by the laser (W/Hz).

Optical Sinusoidal Intensity Modulation

This is the equation for optical sinusoidal intensity modulation. It is called intensity modulation because the optical power is proportional to the electrical modulating signal. This is unlike radio frequency (RF) modulation in which the voltage of the carrier signal is proportional to the voltage or current of the modulating signal.

The carrier term found in the equation for RF modulation is usually left out of the optical modulation equation because the photodetector, unlike an RF detector, does not respond instantaneously to the carrier frequency. Rather, the photodetector responds to the optical power averaged over its response time.

DETECTED RF POWER

JPL



$$P_{rf} = \frac{1}{2} \left[10^{-\frac{P_L}{10}} P_o m \eta \right]^2 R_L$$

where, P_o = the average optical power emitted by the laser (W),

$$m = \frac{P_{\max} - P_o}{P_o} = \text{the modulation index,}$$

P_{\max} = the peak optical power emitted by the laser (W),

P_L = the optical loss between the laser and the photodiode (dBo),

η = the quantum efficiency of the photodetector (A/W), and

R_L = the load across the photodetector (Ohms).

Detected RF Power

This is the equation for the RF power resulting from detection of an optical carrier which has been intensity modulated with a sinusoidal RF signal. Notice that the detected RF power is proportional to the square of the optical loss, received optical power, the modulation index, and the photodetector efficiency. Maximizing the RF signal assures the highest signal-to-noise-ratio, therefore, it is desirable to minimize the optical loss between the laser and the photodetector, to transmit as much optical power as is practical, and to use the highest modulation index consistent with the acceptable signal distortion. It is also desirable to use the largest load resistance consistent with the required bandwidth.

NOISE

JPL



- RELATIVE INTENSITY NOISE (RIN)
- SHOT NOISE
- THERMAL NOISE
- RADIO FREQUENCY (RF) CLOSE-TO-CARRIER (CTC) PHASE NOISE
 - INTERFEROMETRIC INDUCED NOISE
 - EXTERNAL REFLECTION INDUCED LASER NOISE
 - CABLE INSTABILITY

Noise

There are a number of noise sources to be considered in the design of analog fiber optic systems. There is the relative intensity noise (RIN) of the laser, shot noise, thermal noise, and RF close-to-carrier (CTC) phase noise.

RF CTC phase noise has several sources, interferometric noise, reflection induced laser noise, and cable instability.

RELATIVE INTENSITY NOISE (RIN)

JPL



$$RIN = 10 \log \left(\frac{\langle |\delta P|^2 \rangle}{P_o^2} \right) \text{ (dBe/Hz)}$$

where,

δP = optical power fluctuations (W/\sqrt{Hz}), and

P_o = the average optical power (W).

Relative Intensity Noise (RIN)

RIN is the ratio of laser contributed optical noise power squared over the average laser optical power squared expressed in dB. Remember that optical power squared is proportional to electrical power, so another way of expressing RIN is the ratio between the detected noise power and the photodiode current squared.

SHOT NOISE

JPL



$$P_{shot} = 2eI_d R_L \quad (W/Hz)$$

where,

e = the electron charge (C),

I_d = the photocurrent induced by optical power incident on the photodiode (A), and

R_L = the load across the photodiode (Ohms).

Shot Noise

Shot noise is caused by the natural granularity of all light sources with the exception of squeezed light sources. It is beyond the scope of this tutorial to go into the subject of squeezed light other than to say that shot noise amplitude variations can be converted to frequency modulation. When squeezed light is detected with an amplitude detector, the shot noise is reduced.

The signal-to-shot-noise-ratio can be maximized by using the highest possible optical power and modulation index. Shot noise increases linearly with optical power but detected RF signal power increases as the square of the optical power, therefore, the signal-to-shot-noise-ratio increases linearly with optical power. Also, the detected RF signal increases as the square of increasing modulation index but since the average optical power remains constant with increasing modulation index the shot noise remains constant.

THERMAL NOISE

JPL



$$P_{th} = kTF \quad (W/Hz)$$

where,

k = Boltzmann's constant (J/K),

T = the temperature (K), and

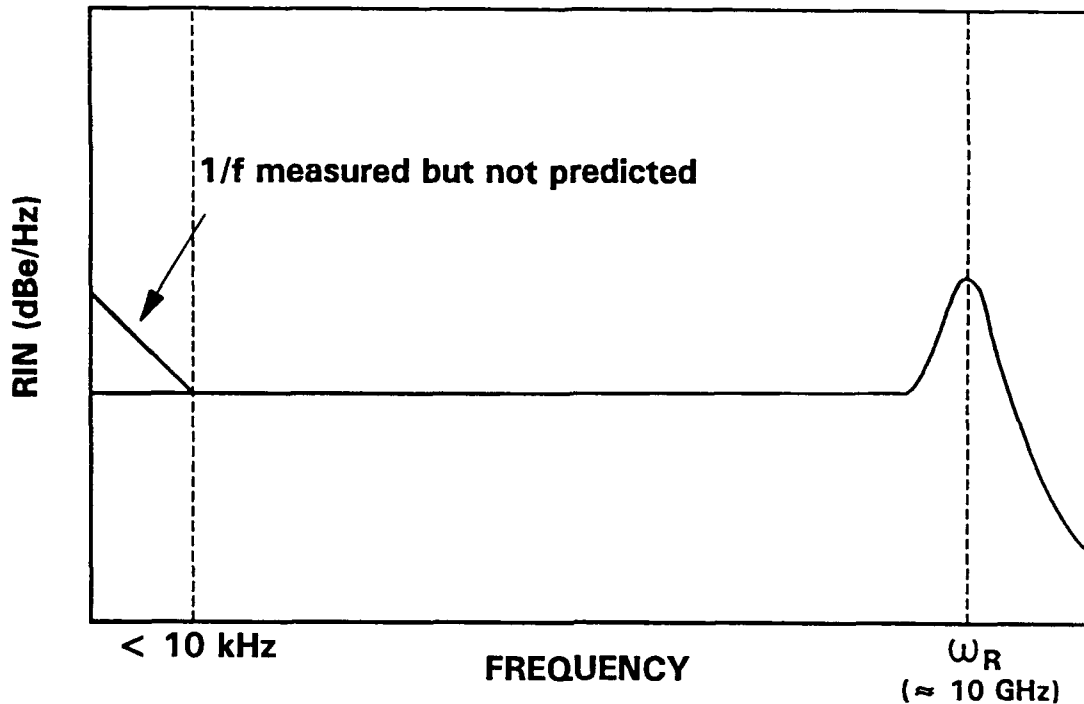
F = the noise figure of the amplifier following the photodiode.

Thermal Noise

In typical analog fiber optic links, for small optical signals the thermal noise of the amplifier following the photodetector predominates. The reason for this is that an amplifier having low input impedance is used after the photodiode to obtain high bandwidth by shunting the photodiode reactance. The low input impedance of this amplifier also shunts the higher impedance photodiode noise source making the amplifier noise predominate.

RIN VS FREQUENCY

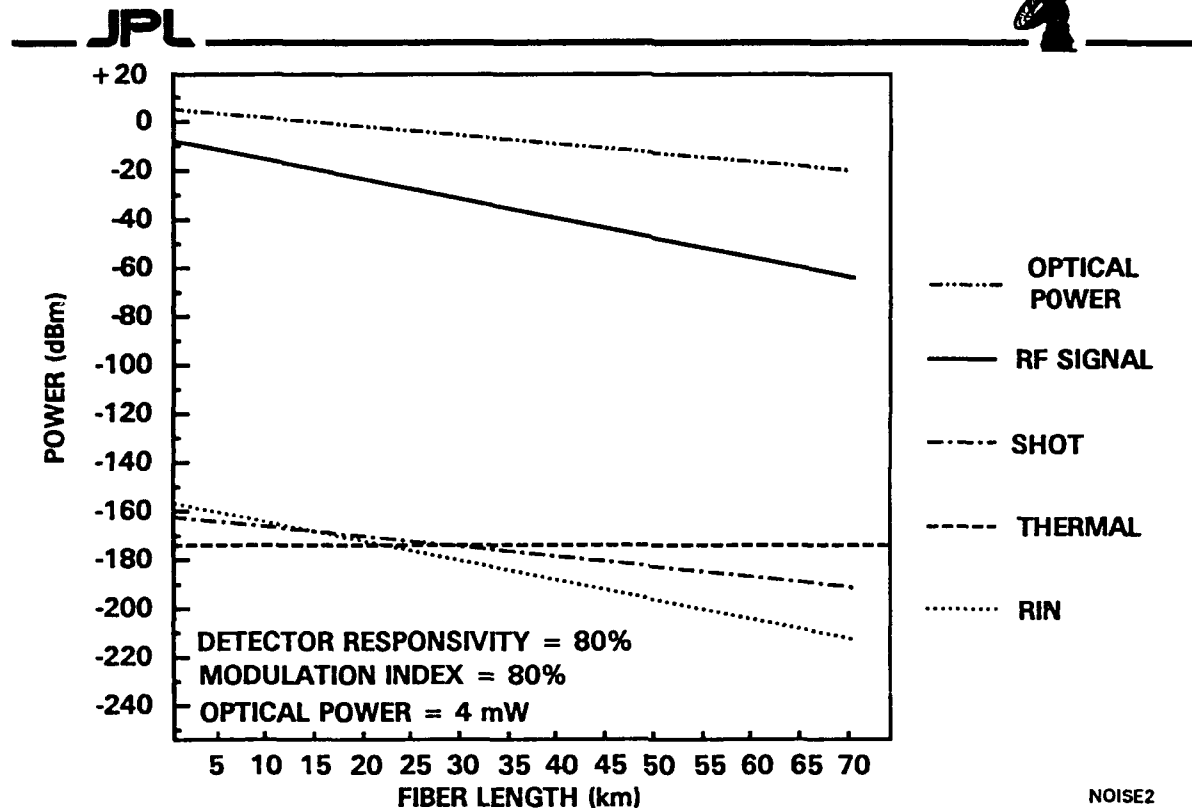
JPL



RIN Versus Frequency

This is a typical curve for relative intensity noise (RIN) for which there is good agreement between theory and measurement. The peak at the high frequency end of the curve is the resonance of the laser cavity. At the low frequency end of the curve $1/f$ noise is shown. One over f noise has several sources, some of which are understood and others which are not predicted by theory.

FIBER OPTIC LINK NOISE SOURCES

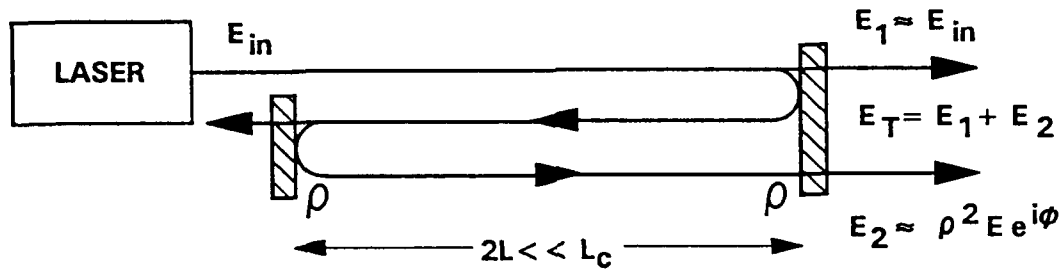


Fiber Optic Link Noise Sources

This is a plot of some of the noise sources in a typical semiconductor laser analog fiber optic link as a function of link loss. It is assumed that the loss in the optical fiber is 0.4 dB/km. Therefore, the optical power decreases at 0.4 dB/km as shown. The detected RF signal decreases as the square of the optical power, 0.8 dB/km. The thermal noise power generated at the receiver input at 23°C is a constant -174 dBm/Hz. The shot noise decreases at the same rate as the optical power and the RIN decreases at the same rate as the detected RF signal. As the optical power decreases, the predominate noise source changes from RIN to shot noise and then to thermal noise.

INTERFEROMETRIC INDUCED NOISE

JPL



$$I \propto E_T^* E_T = I_{\text{avg}} + \Delta I_{\text{pp}} \quad \frac{\Delta I_{\text{pp}}}{I_{\text{avg}}} = \begin{cases} 4\rho^2, & \text{singly polarized} \\ \rho^2, & \text{scrambled} \end{cases}$$

for Fresnel reflection at glass-air: $\rho^2 = 0.04$

Interferometric Induced Noise

Multiple reflections of a signal in a medium result in signals which are traveling in the same direction in the medium. Signals traveling in the same direction in a medium interfere. The signal resulting from this interference is the vector,

$$E_T = E_1 + E_2 = E_{in} (1 + \rho^2 e^{i\phi})$$

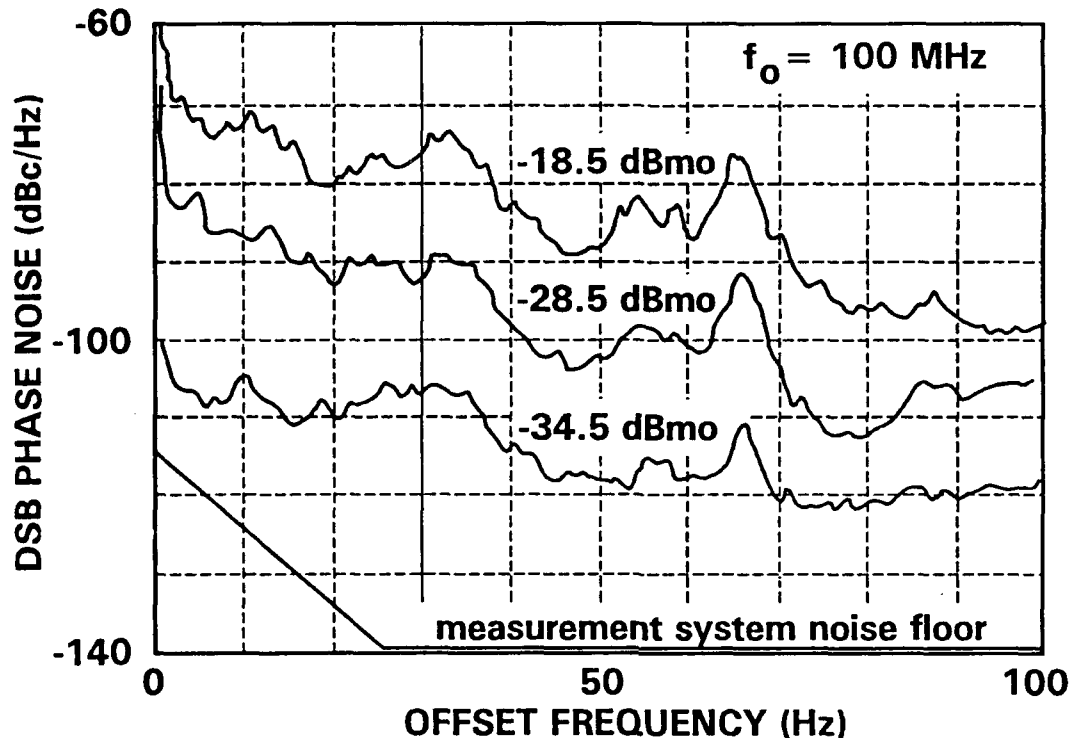
where, $\phi = \frac{2\pi * \eta * \Delta L}{\lambda}$, ΔL = the path difference, and η = the index of refraction of the fiber. Then, $I_T \propto E_T^* E_T$.

The reflected signal travels farther through the medium than the primary signal travels since part of the reflected signal's path is in the opposite direction as shown. Therefore, when the length of the medium changes the relative phase between the primary and reflected signals changes. This change in phase between the two signals, because of interference, results in an unintended amplitude and phase change of the transmitted signal.

At optical wavelengths, 1 to 2 microns, very small changes in relative pathlengths between the reflected and primary signals result in large interferometric effects. When these pathlength changes are generated by environmental variations the resultant noise has a $1/f$ characteristic.

PHASE NOISE VS REFLECTED OPTICAL POWER

JPL



Phase Noise vs Reflected Optical Power

Optical power reflected back into a laser results in increased phase noise on intensity modulation of the optical carrier. The mechanism for this noise, which has a $1/f$ characteristic, is not fully understood.

This graph shows a fiber optic links phase noise contribution to a 100 MHz RF signal which was transmitted through it. The phase noise is plotted as a function of offset frequency from the RF signal and the reflected optical power.

The measured fiber optic link consisted of a directly modulated semiconductor laser with an internal 30 dB optical isolator. Because of this isolator the reflected optical power actually reaching the laser is 30 dB lower than the plotted values. These curves show that the phase noise decreases faster than the reflected power into the laser is decreased.

STIMULATED BRILLOUIN SCATTERING (SBS)

The maximum narrow linewidth optical power which can be transmitted is,

$$P_{CW} = 21 \frac{A_e K}{g_B L_e}$$

where,

$$L_e = \frac{1 - e^{-\alpha \ell}}{\alpha} = \text{the effective interaction length (m),}$$

A_e = the effective core area of the fiber (m^2),

α = the fiber loss per meter (dB/m),

ℓ = the length of the fiber (m),

K = the polarization factor ($1 \leq K \leq 2$), and

g_B = the peak Brillouin gain coefficient ($4.6 \times 10^{-11} \text{ m/W}$).

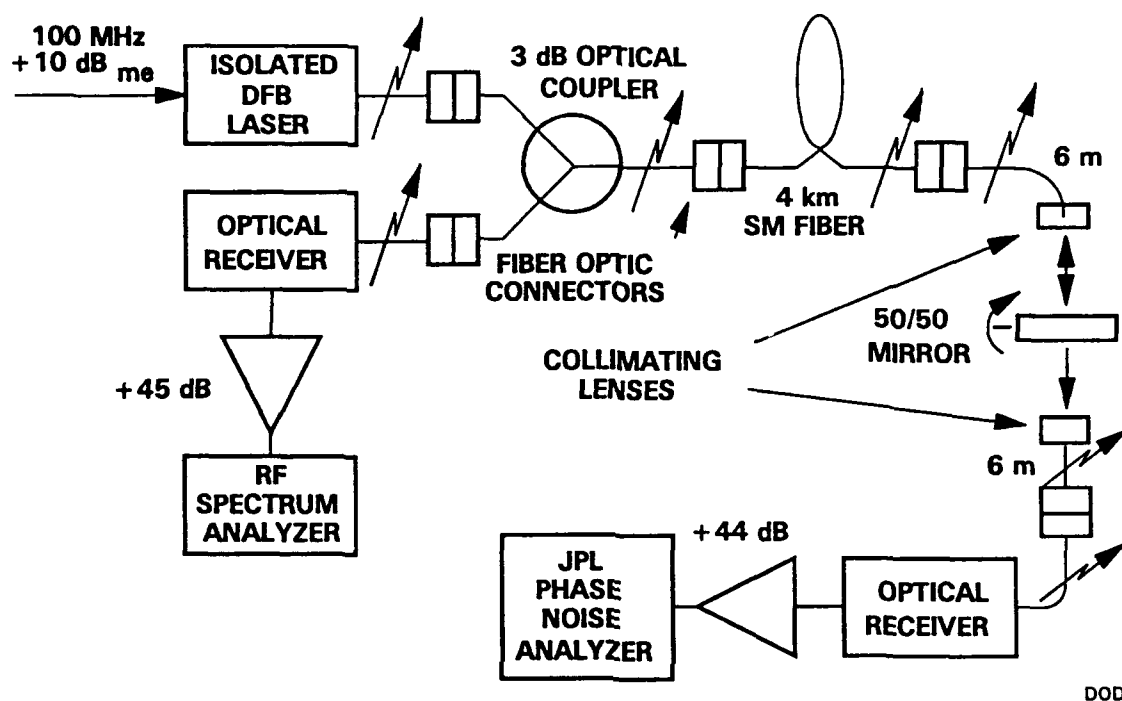
Stimulated Brillouin Scattering

For lasers with less than ≈ 10 MHz linewidth, an effect known as stimulated Brillouin scattering (SBS) limits the maximum optical power which can be transmitted through an optical fiber. Therefore, the SNR of the link will ultimately be limited by SBS. When this optical power level is exceeded additional power is converted to a signal propagating back toward the source. This signal is generated acoustically and is offset in wavelength by about 10 to 15 GHz. Any reflection of this reverse signal will propagate in the same direction as the original signal and interfere with it. This interference will result in intensity modulation at the offset frequency, 10 to 15 GHz, and will show up in the detected output.

MEASUREMENT SYSTEM

PHASE NOISE VS REFLECTED POWER

JPL

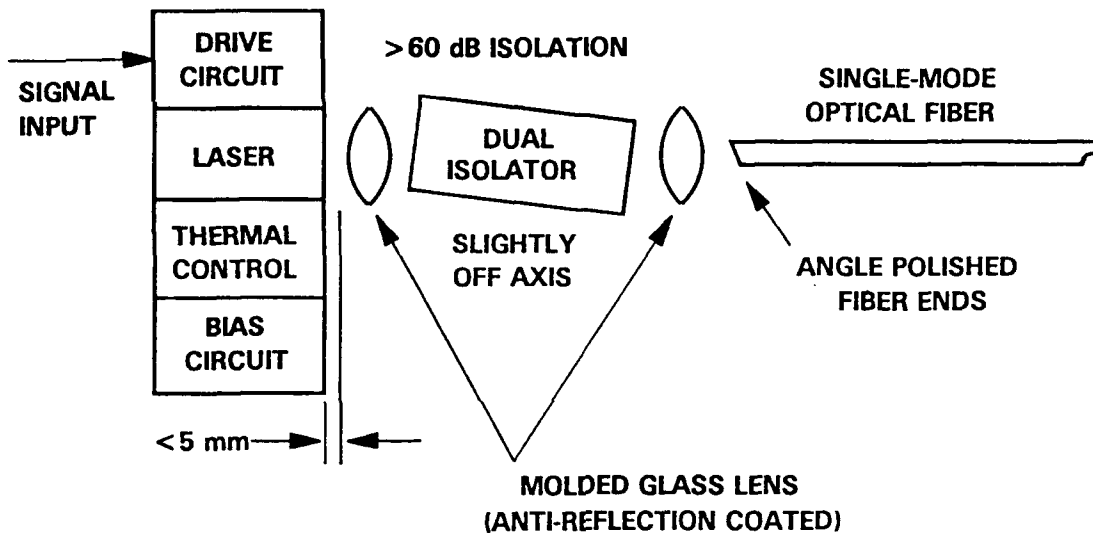


Measurement System

This is the measurement system used to measure the increased phase noise as a result of optical reflections back into a semiconductor laser. The amount of reflected light is varied by adjusting the position of the 50/50 mirror. The magnitude of the reflected light is measured by the RF spectrum analyzer. The RF phase noise of the detected signal is measured at the output of the straight through path by a JPL phase noise analyzer.

LASER PACKAGING FOR IMPROVED CLOSE-TO-CARRIER PHASE NOISE

JPL



Laser Packaging for Improved Close-to-Carrier Phase Noise

This is a block diagram of a suggested coupling system to reduce reflections back into a semiconductor laser. As pointed out earlier these reflections will increase the phase noise of an RF signal which is transmitted through a fiber optic link.

The first reflection should be as close to the laser facet as practical. All surfaces should be antireflection coated and where possible offset from parallel. The ends of the fiber should be slant polished at a 7 or 8 degree angle.

SLOPE OF CABLE GROUP DELAY VS TEMPERATURE

JPL



$$\frac{dD_g}{dT} = \frac{\ell K_{TCD}}{10^6 V_g}$$

where,

K_{TCD} = the thermal coefficient of delay, in ppm/°C, for the cable at the operating temperature,

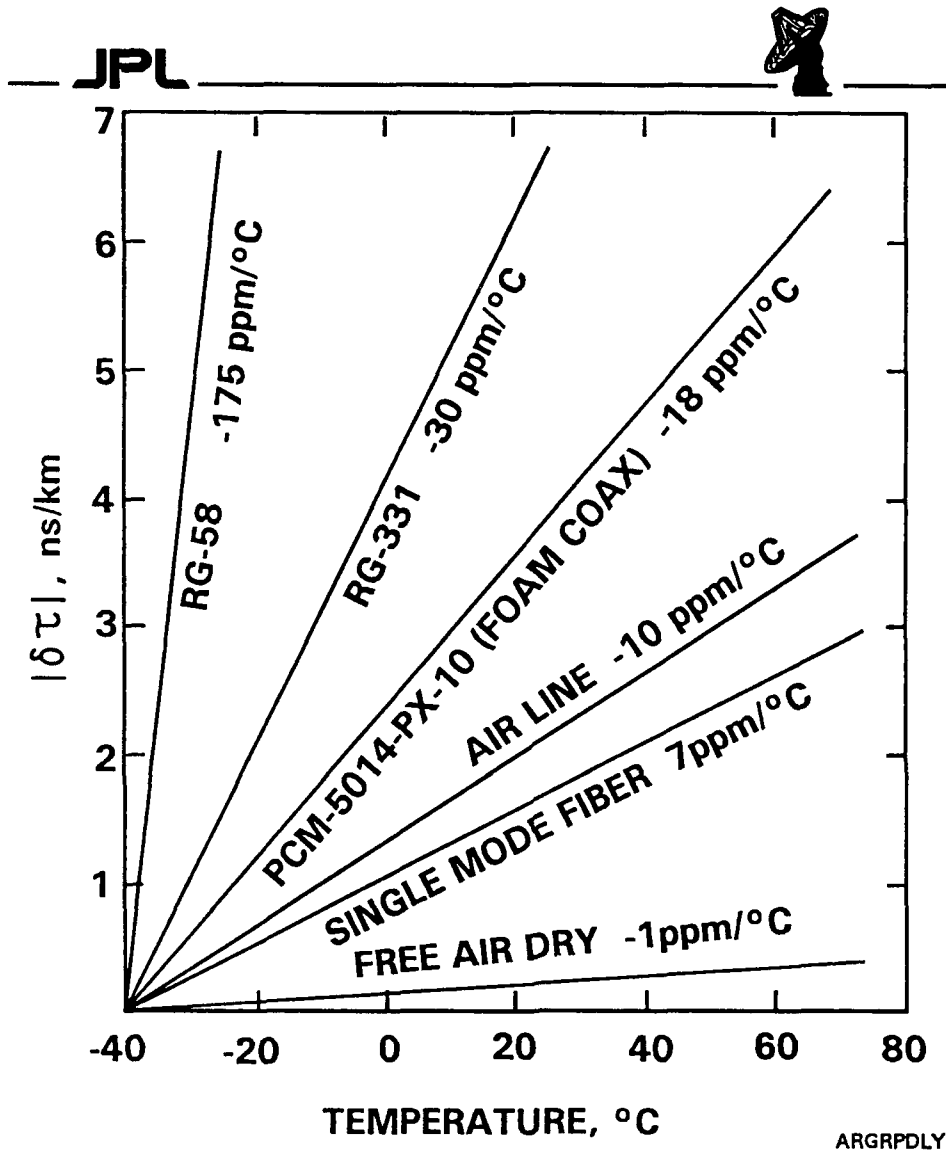
ℓ = the length of the cable (m), and

V_g = the group velocity in the cable (m/s).

Slope of Cable Group Delay vs Temperature

Temperature changes result in group delay changes in all cables. Therefore, when a cable is subjected to changing temperature with respect to time the group delay through the cable changes with time. If the rate of change in group delay versus time is not constant it adds phase noise to a signal which is transmitted through the cable. It is, therefore, desirable to use a cable with the lowest possible slope of group delay vs temperature.

GROUP DELAY VS TEMPERATURE

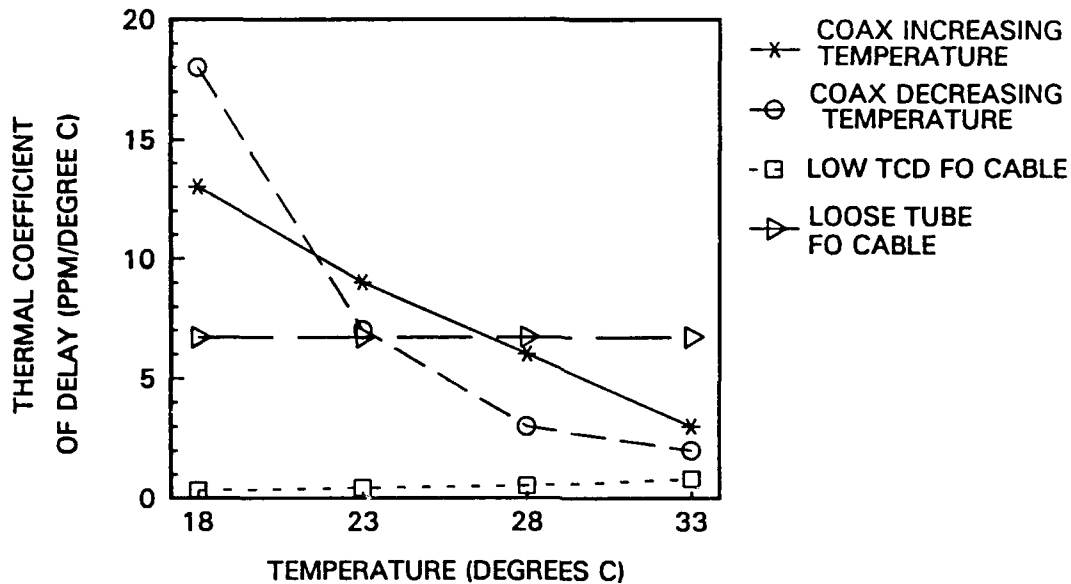


Group Delay vs Temperature

This is a plot of group delay change versus temperature for various types of cable. The change in group delay is given in both nanoseconds per kilometer and parts-per-million (ppm) per degree Celsius. Standard single-mode fiber is $\approx 7 \text{ ppm}/^\circ\text{C}$ and is better than any coaxial cable.

**MEASURED THERMAL COEFFICIENT OF DELAY
FIBER OPTIC CABLE VS BEST COAX (64-875)**

JPL

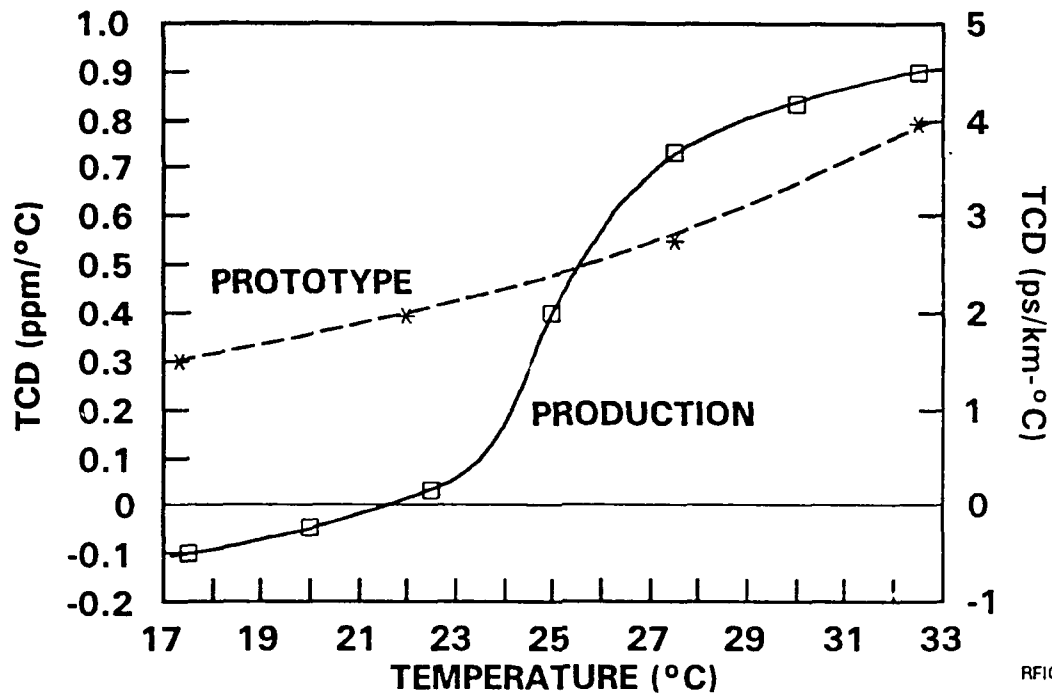


Measured Thermal Coefficient of Delay for Fiber Optic Cable vs Good Coax (64-875)

This is a plot of the thermal coefficient of delay for standard loose tube fiber optic cable and low thermal coefficient of delay fiber optic cable versus high quality 7/8 th inch diameter air dielectric coaxial cable. The slope of group delay versus temperature for the low thermal coefficient of delay fiber optic cable is much smaller than for other cables.

THERMAL COEFFICIENT OF DELAY SUMITOMO FIBER

JPL



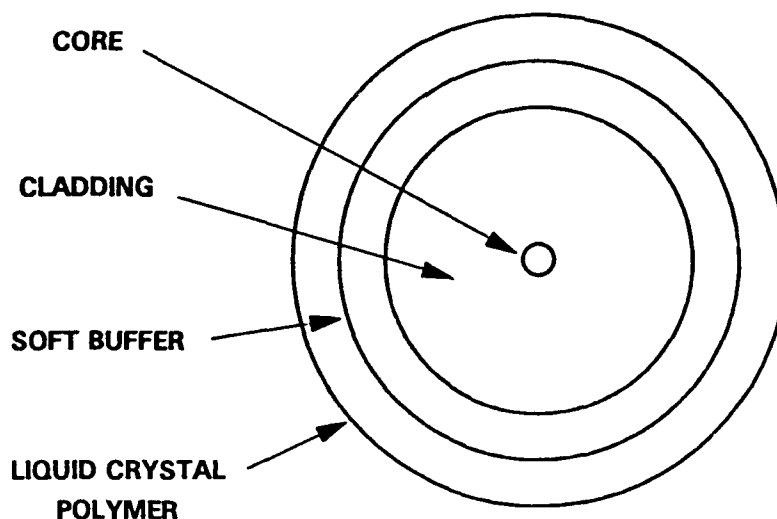
RFIG1

Low Thermal Coefficient of Delay Sumitomo Fiber

This is an expanded plot of the thermal coefficient of delay (TCD) of the low TCD cable. This type of cable is manufactured only by Sumitomo Electric.

LOW TCD OPTICAL FIBER

JPL

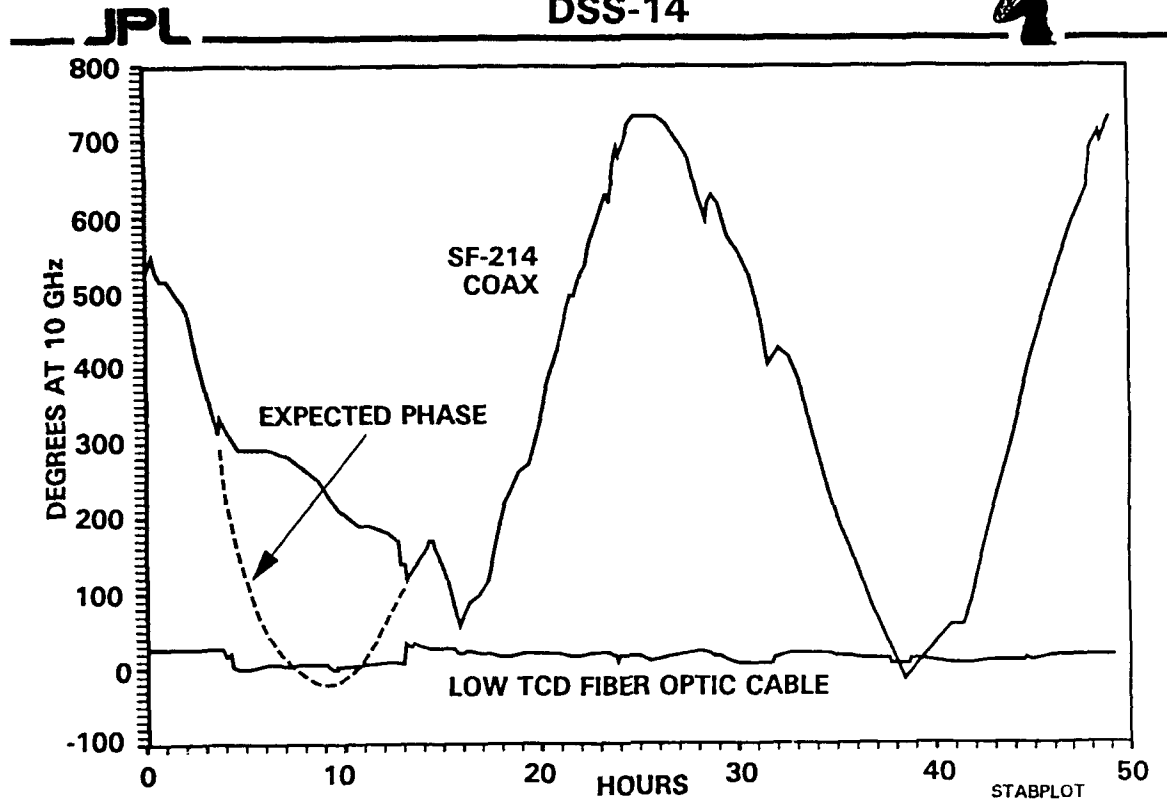


Low TCD Optical Fiber

This is a cross section of low TCD optical fiber. Its low TCD is a result of a liquid crystal coating on the fiber which has a negative thermal coefficient of expansion which opposes the positive coefficient of expansion of the glass fiber. The stress on the fiber which results from these opposing forces restricts the change in length of the fiber and changes its index of refraction resulting in reduction of its thermal coefficient of delay.

The thermal coefficient of expansion of the liquid crystal is higher than that of glass so it would overcompensate if it were coated directly on the fiber. In order to apply the proper amount of stress on the fiber for good compensation a soft stress absorbing buffer is placed between the liquid crystal coating and the fiber. Actually several alternating layers of liquid crystal material and buffer material are used to obtain the proper compensation.

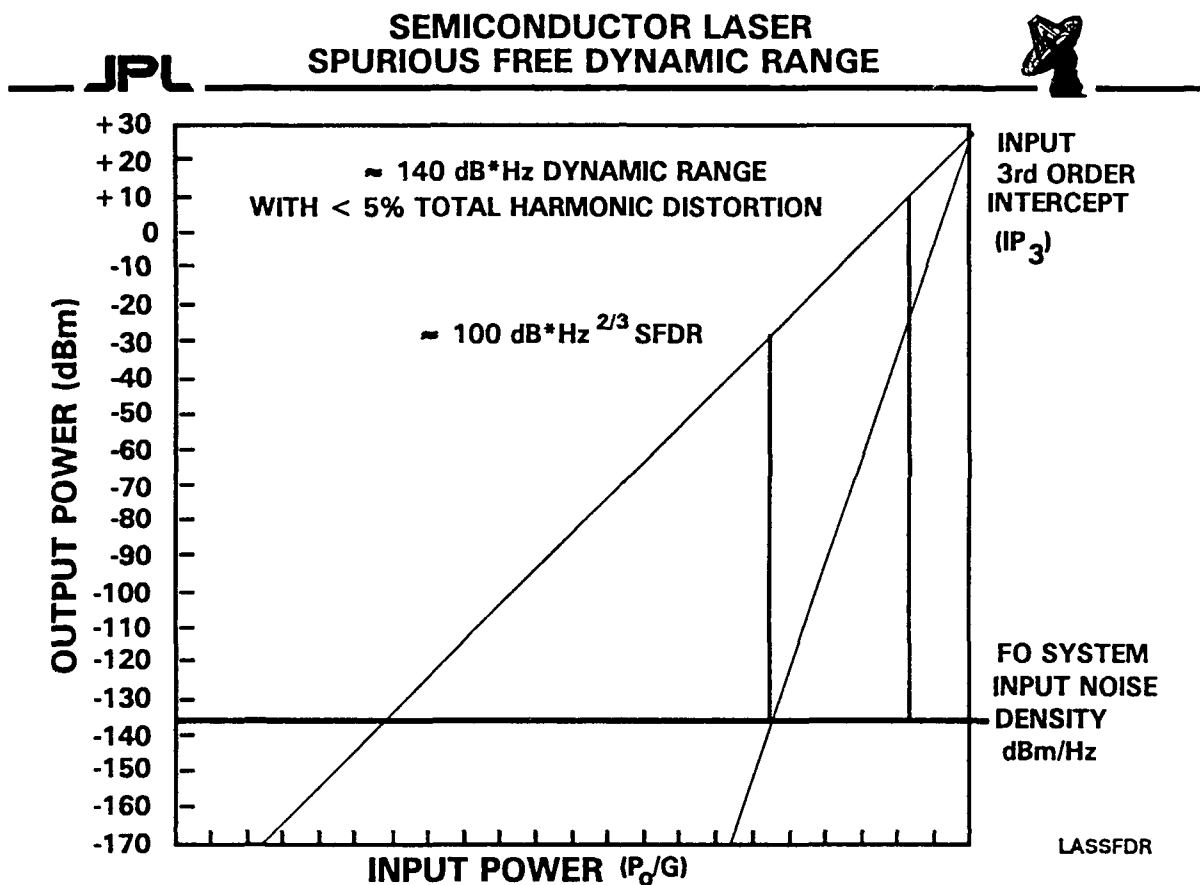
COAX VS FIBER OPTIC CABLE STABILITY DSS-14



Coax vs Fiber Optic Cable Stability

This is a plot of the phase, normalized to 10 GHz, versus time for an SF-214 coaxial cable and a low TCD optical fiber cable which are located in the same antenna wrap-up. The diurnal variations in phase for the coax cable are huge compared to the fiber optic cable. A change in antenna position four hours into the test created stress on both cables which shows up as a variation of phase from the expected norm. The phase of the fiber optic cable changed about 30 degrees and the phase in the coaxial cable changed by about 300 degrees.

The amount of stress generated by antenna movement is related to the configuration of the cable wrap-up in the antenna. The wrap-up configuration is far from optimum in most antennas today because of the large number of cables which must be routed to the antenna cone. When fiber optics is fully implemented in antennas there may be only two cables, one electrical power cable and one fiber optic cable. Much better wrap-up configurations could be used in antennas if the number of cables in the wrap-up was drastically reduced. This could result in virtually no stress related phase changes in the signal transmission cable.



Semiconductor Laser Spurious Free Dynamic Range

This is a plot of the spurious free dynamic range of a fiber optic CATV transmission system manufactured by Ortel Corporation, Alhambra, California. It uses a directly modulated semiconductor laser. Its spurious free dynamic range is approximately $100 \text{ dB*Hz}^{2/3}$ and its dynamic range for less than 5% total harmonic distortion is approximately 140 dB*Hz .

CONDITIONS FOR BEST PERFORMANCE

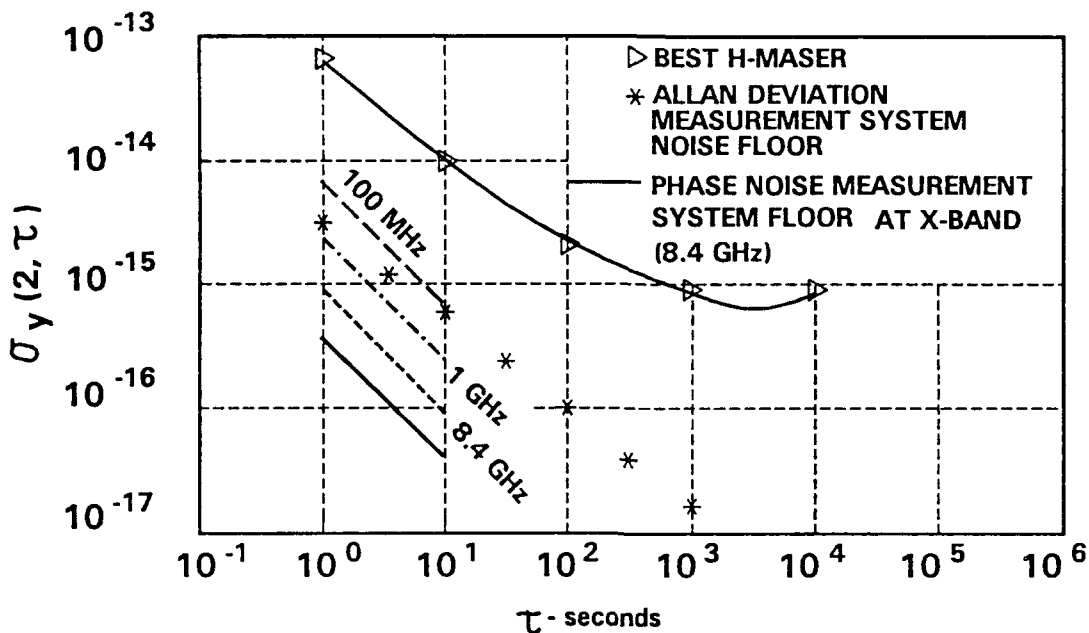
JPL



- TO ACHIEVE BEST PERFORMANCE:
 - DESIGN FOR HIGHEST PRACTICAL SIGNAL-TO-NOISE-RATIO
 - KEEP ALL REFLECTIONS TO A MINIMUM INCLUDING STIMULATED BRILLOUIN SCATTERING
 - USE LOW TCD FIBER AND CONTROL THE ENVIRONMENT
 - USE LASERS WITH INTERNAL OPTICAL ISOLATION AND ADD ADDITIONAL EXTERNAL ISOLATION
 - USE HIGH PERFORMANCE POWER SUPPLY FILTERS WHICH WORK DOWN TO DC

FIBER OPTIC SYSTEM DIFFERENTIAL STABILITY STATUS 1991

JPL

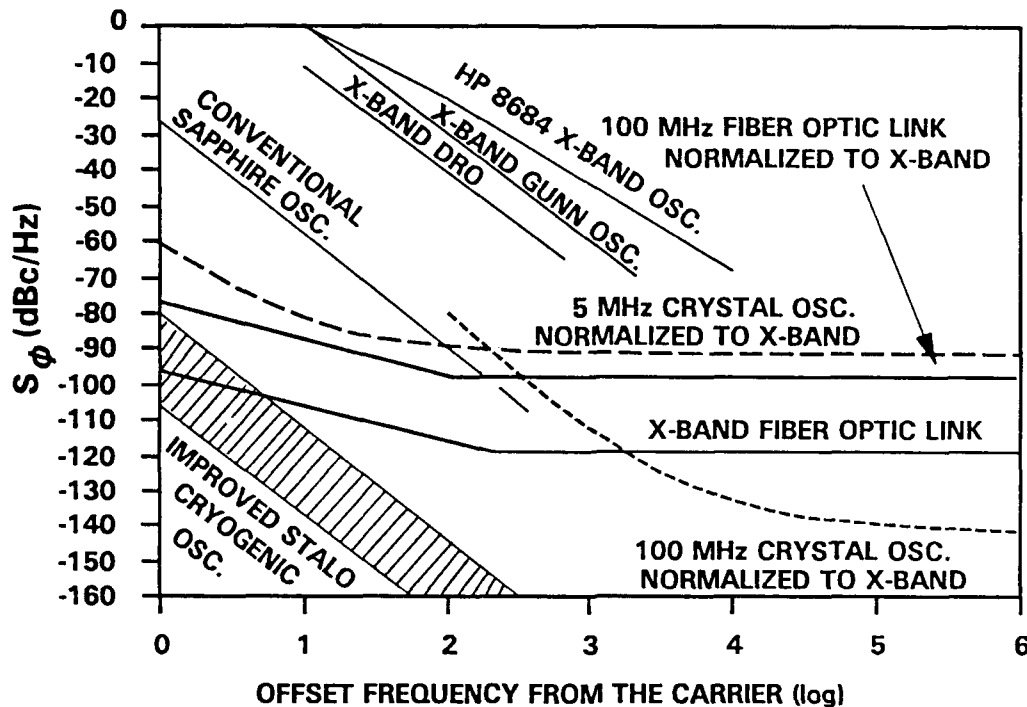


Fiber Optic System Differential Stability Status 1991

This is a plot of the differential Allan deviation, the deviation from one end of a fiber optic link to the other, for fiber optic transmission systems up to the date of this presentation. It shows measured differential Allan deviations for fiber optic transmission at 100 MHz, 1 GHz, and 8.4 GHz compared to the Allan deviation of a typical H-maser.

PHASE NOISE AT X-BAND (1991)

JPL

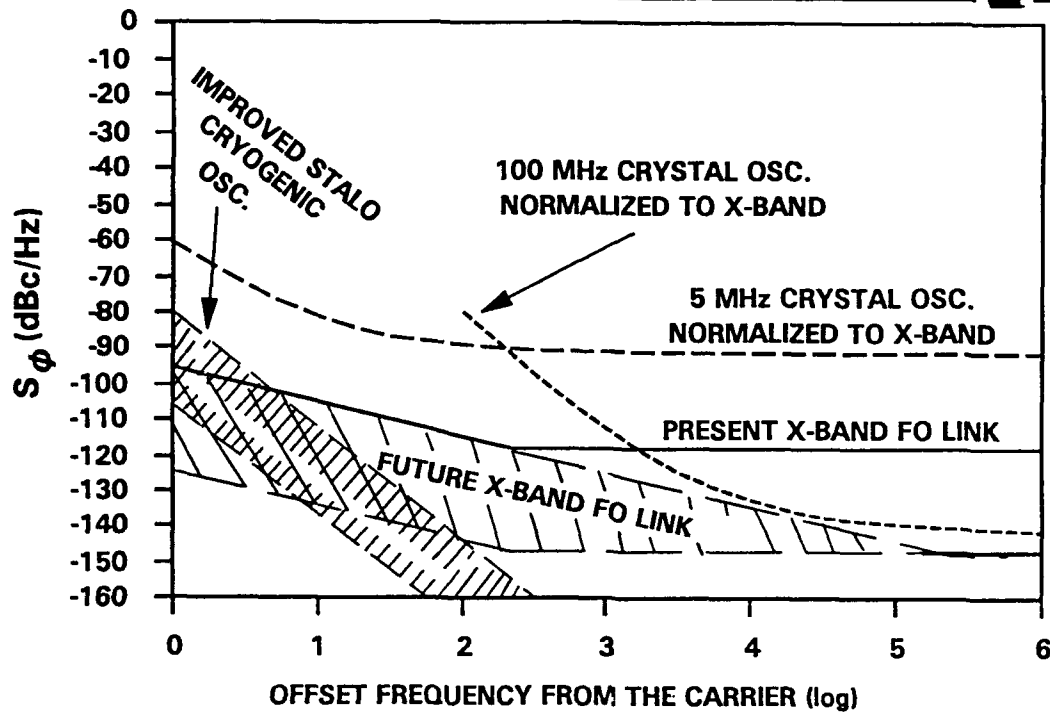


Phase Noise At X-Band (1991)

This is a plot of added phase noise versus offset frequency from the carrier for a 100 MHz and 10 GHz transmission system compared to various signal sources. The phase noise of the 100 MHz transmission system was measured at 100 MHz and normalized to 10 GHz. The phase noise of the sources were either measured at 10 GHz or normalized to 10 GHz as indicated.

PREDICTED FUTURE PHASE NOISE AT X-BAND

JPL



PLOTPHS2

Predicted Future Phase Noise At X-Band

This is a plot of predicted added phase noise for advanced fiber optic transmission systems presently being developed. This predicted phase noise is compared to the best crystal oscillators normalized to 10 GHz and to the STALO cryogenic oscillator.

SUMMARY

JPL



- THIS TUTORIAL HAS:
 - DESCRIBED THE FUNDAMENTAL OPERATION OF TODAY'S ANALOG FO SYSTEMS
 - DISCUSSED THE MAJOR NOISE SOURCES THAT LIMIT THE PERFORMANCE OF ANALOG FO SYSTEMS
 - SHOWN HOW TO ACHIEVE THE BEST PERFORMANCE
 - GIVEN PRESENT AND FUTURE PERFORMANCE OF ANALOG FO SYSTEMS

Introduction to the Time Domain Characterization of Frequency Standards*

James Jespersen
Time & Frequency Division
National Institute of Standards & Technology
Boulder, Colorado 80303

Abstract

The simple mean, the standard deviation and its square, the variance are well known statistical measures of a set of data points. It seems natural, therefore, to apply these measures to signals generated by frequency standards. However, this approach quickly reveals some significant problems.

Normally, when we compute the mean and standard deviation of some process, we assume that including more data points in the computation brings us ever closer to the true mean and standard deviation of the process. Unfortunately this is not true, in general, for data obtained from signals generated by frequency standards. Here, the situation usually deteriorates as more data points are included.

This tutorial describes, largely in heuristic terms, modified statistical measures which do not "blow up" for the kinds of noise processes normally encountered in signals generated by frequency standards.

I. Introduction

This paper is about the question "How do we characterize the performance of frequency standards?" Historically two approaches have developed: one in the frequency domain and the other in the time domain. Our emphasis here will be on the time domain approach; however, we shall begin by considering the frequency domain approach because familiarity with this technique helps to understand why the two usual measures in the time domain, the standard deviation and its square, the variance, don't work when applied to frequency standards.

Let's begin by considering an ideal frequency standard which, of course, does not exist in nature. Such a standard generates a noise-free output with constant amplitude, frequency and phase. Real frequency standards fall short in all three categories. However, variations in amplitude with time are, for practical purposes, negligible, so we won't consider that as a problem in the rest of this paper.

Mathematically, then, the output signal, $V(t)$, of our ideal standard is

$$V(t) = V_p \sin(2\pi\nu_0 t) \quad (1)$$

*Contribution of the U.S. Government, not subject to copyright.

where V_p , is the constant amplitude and ν_0 is some constant frequency.

What real frequency standards generate is

$$V = V_p \sin[2\pi\nu_0 t + \phi(t)] \quad (2)$$

where the phase, ϕ , is a noise perturbation of the desired frequency.

The phase noise, ϕ , falls into two categories: noise we can remove and those we can't. If the signal phase changes predictably with time, a constant frequency offset from the desired frequency for example, we can correct the output by appropriate compensating electronic circuitry or on paper. By definition, the unpredictable, or random part of the signal can't be corrected. Here our only avenue for improvement is to use statistical techniques. However, these statistical improvements are in reality based on assumptions—some model or models—which we believe, or at least suspect, and in a last ditch effort hope, underlie the random part of a signal.

All this adds up, often enough, to a kind of catch 22. Consider a clock whose resonator is a swinging pendulum. We suppose that the clock's misbehavior is related to changes in the weather. Specifically our model says that temperature changes alter the length of the pendulum and thus its natural frequency. We want to predict how much the pendulum clock's time departs, on a daily basis, from time kept by a perfect clock over, say, a period of a month. It's easy. All we have to do is foretell the weather day by day for the next month.

Here our model is probably correct but it doesn't help much in any detailed way. The best we can do, in the absence of actual temperature measurements, is depend on local weather forecasts which are notoriously unreliable.

The actual situation for frequency standards may be worse. A good deal of statistical information about frequency standards has accumulated from numerous measurements over the years, but nobody has any respectable models for what causes the most troublesome kinds of noise. Put differently, the models themselves are statistical models based on measurements, not statistical models based on the physics of the underlying behavior. It's as though we have reams of data giving the average temperature but we know nothing whatsoever about the laws governing atmospheric processes. With this state of affairs about the best we can do is say the temperature tomorrow will be about what it was today.

Nevertheless, much has been gained by inspecting frequency standards through the eyes of statisticians. As an example, in section VI, we consider how various statistical models of frequency standard performance can be applied to determining the performance of a clock over a specified time.

II. Measuring Performance in the Frequency Domain

Consider again the output of an ideal frequency standard as expressed by Equation 1. Figure 1a graphically represents Equation 1 in the frequency domain. This figure shows that the output is at frequency ν_0 with amplitude V_p . A variation on Figure 1a, Figure 1b, is to plot the power carried by the signal at frequency ν_0 , rather than the amplitude. This representation is called the power spectrum of the signal, and we shall use this representation from now on for reasons which will be apparent later. We can easily extend the power spectrum representation of a signal, as in Figure

1c, to represent, say, the superposition of three sinusoids at frequencies ν_1 , ν_2 , and ν_3 with powers P1, P2, and P3.

The power spectrum approach is a convenient way to represent a sinusoid which has been perturbed by noise. We can think of the noise as "blurring" the frequency of the sinusoid so that the power spectrum of a real oscillator is more like the one shown in Figure 2 than the idealization in Figure 1b. The frequency spread of the power spectrum curve is a measure of the amount of noise.

The power spectrum in Figure 3 includes both the signal at ν_0 and the noise components at other frequencies, but we are interested primarily in the power spectrum of the oscillator noise. We can isolate this noise as follows. The actual frequency output, $\nu(t)$, of our oscillator signal is just the rate of change in the total phase of the sinusoid given by Equation 2.

$$2\pi\nu(t) = \frac{d}{dt}(2\pi\nu_0 t + \phi(t)) = 2\pi\nu_0 + \frac{d}{dt}(\phi(t)) \quad (3)$$

That is $\nu(t) - \nu_0 = \frac{1}{2\pi} \frac{d}{dt}(\phi(t))$ is the amount by which the output frequency deviates from the desired frequency. A quantity, called the fractional frequency fluctuation, $y(t) = \frac{\nu(t) - \nu_0}{\nu_0}$ is a dimensionless quantity with greater utility than $(\nu(t) - \nu_0)$ as illustrated by the following example:

Suppose the quantity $(\nu(t) - \nu_0)$ equals 1 Hz for two different oscillators. Are the two oscillators equal in quality? Not if one oscillator operates at 10 Hz and the other at 10 MHz. In the first case, the average value of the fractional frequency fluctuation is 1/10, and in the second 1/10,000,000 or 1×10^{-7} . The 10 MHz oscillator is thus more precise—has a higher Q. If frequencies are multiplied or divided, using ideal electronics, the fractional stability is not changed. Since $y(t)$ represents the fractional frequency excursions of the signal frequency about the nominal frequency, ν_0 , the power spectrum of $y(t)$ characterizes the oscillator noise independent of the oscillator signal frequency.

III. Real Oscillator Noise Viewed from the Frequency Domain

Many workers have measured the noise associated with high quality oscillators. Typically the measurements yield the result shown in Figure 3. Here we show the power spectral density, $S_y(f)$ of the fractional frequency fluctuations $y(t)$. The power spectral density $S_y(f)$, is simply the power spectrum normalized so that the area under the $S_y(f)$ curve equals one. $S_y(f)$ is one of the two commonly used measures of frequency instability—the one in the frequency domain. The figure shows for frequencies ranging from 10 Hz to about 0.01 Hz the noise power spectral density is constant with frequency. This kind of noise, called "white noise FM", can be reduced by averaging measurements taken over an extended period of time. The reason it can be "averaged out" is because these noise fluctuations, on the average, advance the phase of the signal as much as they delay it.

In the range of frequencies from about 0.01 Hz to about 10^{-4} Hz, the noise density increases as f^{-1} . This kind of noise is called "flicker noise FM" and its cause in high quality oscillators is not well understood. However, it is probably related to power supply voltage fluctuations, magnetic field fluctuations, and component changes, etc. The frequency range over which it dominates, varies from oscillator to oscillator.

At frequencies lower than 10^{-4} Hz, noise with an f^{-2} dependency is commonly found. This noise, "random walk FM noise", is probably related to environmental effects such as mechanical shock, temperature variations, vibrations, etc. which cause random shifts in the carrier oscillator signal frequency. It can be minimized by carefully isolating the oscillator from its environment.

IV. Effect of Record Length and Sampling Rate on Estimating $S_y(f)$

Suppose we have an analog record of oscillator noise, $y(t)$, 50 seconds long, Figure 4a. We sample this record once every second, producing 50 discrete data points y_1, y_2, y_3 , etc. This record contains a mixture of noise fluctuations at many different rates or Fourier frequencies. However, the sampling rate, once per second, and the length of the record (50 seconds), place bounds on the range of frequencies which can be recovered from the sampled data. First, the sampling rate dictates the highest frequency, the Nyquist frequency, which can be recovered or resolved from the data points, as illustrated in Figure 4b.

Samples are taken at the instants marked by the dots. At this sampling rate we can resolve the fluctuation labeled 1 Hz. Fluctuations at a higher rate, such as the one labeled 5 Hz will be confused with frequency fluctuations at lower frequencies. In this particular example there is no way to distinguish between the signal fluctuations at 1 Hz and at 5 Hz. To avoid this kind of confusion we must either sample at a rate which is high enough to resolve the highest rate of noise fluctuation we expect in the oscillator noise spectrum, or alternately, we must remove all frequencies above the highest frequency which can be resolved by our sampling rate. This could be done by low pass filtering the data before sampling. Later in this section we shall describe a digital filter which accomplishes this purpose; although, the data could be filtered with an analog filter to achieve the same result.

In general the highest frequency we can resolve by sampling every Δt seconds, is $\frac{1}{2\Delta t}$.

The length of the record dictates the lowest noise frequency we can detect. Obviously we cannot detect a fluctuation with a one year period by inspecting a record 50 seconds long. More precisely, the lowest frequency we can detect from a record T seconds long is $2/T$ Hz or .04 Hz when $T = 50$ seconds.

Most procedures developed to determine the power spectral density from a series of discrete data points assume that the sampling rate is several times faster than needed to resolve the highest fluctuation frequency expected in the data and the record several times longer than needed to resolve the lowest frequency expected.

V. Time Domain Characterization of Stability

In the preceding section we showed how oscillator noise can be characterized in the frequency domain. As a practical matter, the output signals are often sampled to obtain a discrete set of y_i and then mathematically manipulated to obtain an estimate of $S_y(f)$. This often requires considerable computation. However, for the types of noises which plague most oscillators, there is a considerably simpler approach (from a computational point of view) which is the second, commonly used, characterization of instability mentioned in the introduction.

In this method we again start with a set of y_i , but instead of converting these measurements to a frequency domain representation we develop a procedure for determining the dispersion or scatter of the y_i as a measure of oscillator noise. The bigger the scatter or dispersion of the y_i the greater the instability of the output signal of the oscillator.

The most common statistical measure of dispersion is the standard deviation (or equivalently the variance which is the square of the standard deviation). The variance, R^2 , is simply a measure of the numerical spread of a set of data points with respect to the average or mean value, \bar{y} , of that set of data points. The formula for computing the variance of a set of M data points taken at uniform intervals of time is:

$$R^2 = \frac{1}{M-1} \sum_{i=1}^M (y_i - \bar{y})^2$$

The reason for squaring the terms $(y_i - \bar{y})$ is that an excursion in y_i from the mean always gives a positive contribution in the calculation of scatter. That is an excursion in the positive direction is not canceled by one in the negative, as is the case for the mean.

A very simple measure of the instability of an oscillator is to simply compute the variance of the y_i obtained from a particular oscillator. We would expect our confidence in our measurement of the variance to increase as we increased the number of data points used in our computations. Unfortunately, for common types of oscillator noise, this is only true to a point. To understand this better refer back to Figure 3 which illustrates the spectral noise density typical of frequency standards. Suppose we have sampled the oscillator noise once per second. We decide to estimate the variance of $y(t)$ from 100 consecutive data points. As we have seen already, data taken at intervals of one second over a record length of 50 seconds effectively "isolates" those noise fluctuations with frequencies in the 0.5 to .04 Hz range. Or stated differently, only those noise fluctuations in the 0.5 to 0.04 frequency range contribute to our estimate of the variance. As we increase the data length, lower and lower noise frequencies contribute to the variance. As we see from Figure 3, however, the noise fluctuations have a constant power down to about 0.01 Hz. This means that as long as our data length does not exceed 200 seconds, the variance does not depend on record length. (It is true, however, that our confidence in the estimate of the variance depends on the number of data points, so that, for best confidence in the variance, we should use 200 data points.)

If we extend our calculations to include data taken over an interval which exceeds 200 seconds, the variance will be influenced by noise fluctuations from that part of the noise spectrum which is not flat with frequency. Now the variance grows in magnitude with data length. This is unfortunate. We would like our statistical measure of scatter to converge toward some value as we increase the number of values used in our estimation—not move off to new values with increasing data length.

We must modify our calculation of variance as applied to oscillator noise, so that the variance converges toward some value as we increase the number of data points employed in our calculations.

VI. Stationarity

In the previous section we saw that the variance does not yield a satisfactory result when applied to frequency standard noise. The problem we are encountering is similar to the problem that arises when we attempt to compute the average value for data which has a slow drift. This drift might be due to some instrumental fault so the drift is an artifact of the measurement procedure.

Consider the record shown in Figure 5-a. These data have random fluctuations superimposed on a slow drift. Obviously the mean value computed for the first half of the record does not equal the mean computed for the second half. In this example the mean depends on when the mean is computed.

With stationary data the mean and variance converge toward some value as the record length increased. Obviously data with a slow drift is non-stationary so the mean and variance never converge to any particular value.

We would like to have a measure of dispersion of frequency standard noise which converges toward some value as record length increases. If we apply the normal variance, R^2 , calculation to the data shown in Figure 5a, this will not be so; since, the mean, which enters into the variance calculation, changes with record length.

Suppose, however, that we calculate another quantity, d^2 , which is similar to R^2 except that instead of subtracting the mean from each data point before squaring their summation, we subtract the previous data point. That is:

$$d^2 = \frac{1}{M-1} \sum_{i=1}^M (y_i - y_{i-1})^2. \quad (4)$$

The terms $(y_i - y_{i-1}) = \Delta y_i$ take on nearly the same values that they would whether a trend is present or not i.e., the quantity d^2 is insensitive to the absence or presence of a trend, while R^2 is sensitive.

Let us examine d^2 from a different point of view. Figure 5b shows a plot of the Δy_i 's. The Δy_i plot is similar to Figure 5a except that the trend is missing. In fact, we can think of the process of taking successive differences (computing the Δy 's) as equivalent to running the data in Figure 5-a through a high pass filter which removes the trend. Equation 3, then, is akin to computing R^2 but contains the additional feature of removing the trend. With the trend removed, we would expect d^2 to converge with increasing data length where R^2 did not.

Let's consider now, in more detail, the nature of the high pass filter produced by taking successive differences. We could determine experimentally the frequency response of this filter by taking a large number of digitized sine waves with the same amplitude but with different frequencies. Suppose for example we took a 1 Hz sinusoid sampled ten times per second to produce a set of y_i 's, Figure 6a. We now take successive differences between the y_i to produce the Δy_i which are also shown in Figure 6b. As we see we obtain, again, a sinusoid of the same frequency but with reduced amplitude (and shifted phase which is not important here). Similarly we could take other sinusoids of unit amplitude and in a similar manner determine their amplitudes upon taking successive differences. The result of such an experiment is shown in Figure 7a. As we see the filter does not attenuate the frequency at the reciprocal of two times the sampling rate (0.05 Hz) but does attenuate all lower frequencies with the attenuation increasing as the frequency approaches zero.

In some situations we might want to leave the trend in and remove the fast fluctuations. For example, suppose we have data consisting of hourly temperature readings taken over a period of one year. If we average the hourly temperature readings, a day at a time, we remove the hourly fluctuations leaving just the day to day trend. The response of this low-pass averaging filter is shown in Figure 7b. It has unity gain at 0 Hz and zero gain at $1/2$ (averaging interval) $= 1/[2 (24 \text{ hours})] = 5.8 \times 10^{-6} \text{ Hz}$.

Suppose, now, we have a set of measurements which we first average, corresponding to low-pass filtering, and then difference, corresponding to high-pass filtering. We can think of these two operations together as equivalent to band pass filtering the original data. The shape of this band pass filter is simply the product of the low and high pass filters, Figure 8a.

Figure 8a shows the response shape of the band pass filter corresponding to first averaging the data two seconds at a time and then taking successive differences. As we increase the time over which we average the data and then take differences, the frequency at which the band pass filter has maximum response moves toward a lower frequency and the width of the filter decreases. To demonstrate this we have plotted in Figure 8-b the band pass filter response corresponding to averaging and differencing times of 2, 4, and 10 seconds.

From this illustration we see that by adjusting the length of the averaging time of the original data points, we can move the center frequency of the filter. This suggests that we might use such a band pass filter to estimate the power spectral density of oscillator noise. Suppose we first average and difference our data, the y_i , with the filter centered at 1.0 Hz. The filtered data now contains only those noise components near 1 Hz. From this filtered data we compute a value for the variance. Next we center the filter at 0.5 Hz and again compute the variance from the filtered data. Now we have computed the variance for the noise fluctuations in the vicinity of 0.5 Hz. We can continue this process indefinitely letting the filter center frequency approach 0.0 Hz. If we now plot the variances obtained as a function of the associated center frequency of the filter, we obtain an estimate of the noise power as a function of frequency which is similar to the kind of information contained in a power spectral density, $S_y(f)$, plot. This is the reason we have chosen to work with the variance rather than the standard deviation since the standard deviation has the dimensions of amplitude.

It can be shown, on a rigorous mathematical basis, that if we pass noise with a $1/f$ or $1/f^2$ power spectral dependence through the type of digital filter we have been discussing, that the noise passing through this filter is stationary. If we now compute the variance of this filtered noise, we can be assured that the variance will converge as the record length increases. Further this variance is a measure of the noise fluctuations in a particular frequency band which is dependent upon the frequency at which the bandpass filter is centered.

VII. Pair Variance

With the background developed in the previous section, we can now define explicitly the second measure of frequency instability (the one in the time domain) which is commonly employed. This measure, often called the "pair variance," is defined by:

$$\sigma_y^2(\tau) = \frac{1}{2(M-1)} \sum_{i=1}^M (y_i - y_{i-1})^2$$

where M is the number of data points and the y_i are obtained by averaging the data, in M segments, each τ seconds long.

We can consider this formula in the light of the bandpass filtering discussed in the previous section. First, we notice that the y_i are obtained by averaging the signal in chunks τ seconds long, which corresponds to low-pass filtering the data. Second, the differences are taken between successive pairs of data which are separated in time by τ seconds; because, the original data was averaged τ

seconds at a time to create the y_i . As we know the averaging and differencing procedures amount to band pass filtering the data to produce stationary noise—assuming the spectral noise density varies as $1/f$ or $1/f^2$. We notice that there is a factor $1/[2(M-1)]$ in front of the summation sign which differs by a factor of $1/2$ from the formula for R^2 . This is because we want the two formulas for R^2 and σ_y^2 to give the same numerical result for white noise FM—which is stationary.

Let's apply the variance to oscillation noise of the type shown in Figure 3. Suppose we compute $\sigma_y(\tau)$, the square root of the pair variance, for $\tau = 10$ seconds. This corresponds to calculating the standard deviation for data which has passed through the filter labeled $\tau = 10$ seconds in Figure 8b. As we increase τ , our band pass filter moves toward lower frequencies and narrows. As long as the spectrum is constant with frequency, σ_y decreases with τ ; since, the filter narrows as it moves toward lower frequencies. As we move into the f^{-1} portion of the spectrum, the bandpass filter continues to narrow as the noise power increases. In this region the filter narrows at a rate which exactly compensates for the increase in noise power so that σ_y is flat in this range. This flat region is called the "flicker floor" of the oscillator, and from a practical point of view, once the flicker floor is reached, no further gains are made by averaging. As we move into the f^{-2} region, the noise power increases at a rate which is faster than is compensated for by the narrowing filter so that in this region σ_y grows with τ .

Obviously the variation of $\sigma_y(\tau)$ with τ is related to the spectral density noise $S_y(f)$ as indicated by the arrows connecting the two parts of Figure 9. In fact, it can be shown mathematically that the slopes of the σ_y vs τ curve are -1 , 0 , and $+1$ for f^0 , f^{-1} , and f^{-2} types of noise respectively. We see then that the σ_y vs (τ) curve is a fairly simple way, from a computational point of view, to infer the noise spectrum of the oscillator.

Although we have been particularly concerned with f^{-1} and f^{-2} noise, it can be shown as we said that the pair variance converges for noise which is divergent up to, but not including f^{-3} noise. At f^{-3} , and beyond, $\sigma_y(\tau)$ does not converge as the data length increases.

We have considered two measures of frequency stability: one in the frequency domain, $S_y(f)$ and the other in the time domain, $\sigma_y^2(\tau)$. From a computational point of view, σ_y^2 is a convenient measure of stability and can be used to characterize the spectrum of oscillation noise. On the other hand, the power spectral density, $S_y(f)$, is quite often more convenient for theoretical developments since it is a fundamental quality. Cutler and Searle (reference 1) have shown that σ_y^2 and $S_y(f)$ are related as follows:

$$\sigma_y(\tau) = \int_{-\infty}^{\infty} df S_y(f) \frac{\sin^4 f}{(\pi f \tau)^2}$$

This equation is the analog of what we have developed from a digital filter point of view. The term:

$$\frac{\sin^4 f}{(\pi f \tau)^2}$$

can be thought of as a band pass filter, through which the noise spectrum $S_y(f)$ is passed, during the integration over f , to determine σ_y for a particular τ .

VIII. Characteristics of Real Frequency Standards and Clocks

Now that we are armed with the primary time domain performance measure of frequency standards let's see how the various kinds of standards stack up. Figure 10 illustrates what we might expect from a number of different devices.

For a cesium frequency standard we have displayed three curves: one represents a standard portable commercial device, another a high performance commercial standard, and finally, a laboratory standard which would be found normally at a national standards laboratory. We include both a standard and a high performance rubidium standard. OCXO means an oven controlled crystal oscillator.

With these characterizations let's consider one example of how we might use these performance measures: the problem of predicting the performance of clocks whose resonators are based on the frequency standard characteristics discussed in the previous section.

By now we know that the kind of noise that dominates, for a particular frequency standard, depends on the sample time τ . For example we see, from Figure 3 that a standard rubidium device is dominated with flicker noise FM—the flat part of any of the curves in the figure—at about $\tau = 1000$ seconds while the transition for the passive hydrogen maser doesn't occur until about $\tau = 100,000$ seconds.

Since a clock is based on a frequency standard, the degree to which a clock departs from a perfect clock over time depends on the kind of noise dominating the frequency standard for the prediction time of interest. In our discussion we assume that constant frequency offsets and linear drifts in frequency have been removed from the frequency standard driving our clock, and that only the random noise is keeping our clock from perfection.

For the three regimes of noise illustrated in Figure 3 reference 2 shows that:

1. When white noise FM and random walk noise predominate, the optimum predictor for clock dispersion is simply $\tau\sigma_y(\tau)$
2. when flicker noise FM is dominant the optimum predictor is $\tau\sigma_y(\tau)\ln 2$.

Figure 11 shows, applying these rules, the time dispersion—the RMS time prediction error—we would expect for active and passive masers, a cesium standard, and a rubidium standard.

As a specific example, consider a large collection of clocks all of whose resonators are cesium frequency standards with the same noise properties. We want to know the RMS scatter of these clocks after, say, 100,000 seconds—a little over one day—after they have been synchronized. As Figure 11 shows at 100,000 seconds the RMS spread is about 5 nanoseconds. From a practical point of view this means that if we have a large collection of clocks, and want to keep them synchronized to about 5 nanosecond, we need to resynchronize them about once a day.

If these clocks were at nodes in a communications systems we would need to dedicate some of our communication channel to achieve this level of synchronization, or alternately recalibrate against some external time source, say a satellite time signal, about once a day.

Of course clocks based on other kinds of resonators would need resynchronization more or less often depending on the quality of the resonator.

IX. Concluding Remarks

The notion of averaging and differencing successive data points to remove both short term and long term variations is an old one. Meteorologists have used these techniques for at least 150 years in their studies of temperature and pressure variations. In more recent times a 1942 paper by von Neuman *et. al.* explicitly states the problem of applying the classical standard deviation to data containing a trend [reference 3]:

"There are cases, however, where the standard deviation may be held constant, but the mean varies from one observation to the next. If no correction is made for such variation of the mean, and the standard deviation is computed from the data in the conventional way, then the estimated standard deviation will tend to be larger than the true population value. When the variation in the mean is gradual, so that a trend (which need not be linear) is shifting the mean of the population, a rather simple method of minimizing the effect of the trend on dispersion is to estimate standard deviation from differences."

Clock metrologists have developed these ideas into rather sophisticated but relatively simple tools for the study of the important noise processes in frequency standards. In the time and frequency literature you will find the pair variance also referred to as the "two sample variance" or the "Allan variance" after David Allan [reference 4] who has focused these techniques on frequency standards. But the problem of characterizing oscillator frequency stability is by no means closed. We have discussed the most commonly encountered measures in the frequency and time domains, but anyone who wishes to pursue current investigations will find reference 5 a good place to start.

References

1. Cutler, L. and Searle, C., *"Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards."* Proc. IEEE, Vol. 54, pp. 136-154, Feb., 1966
2. Allan, D. and Hellwig, H., *"Time Deviation and Time Prediction Error for Clock Specification, Characterization, and Application."* PLANS 1978 Conf. Proc., pp. 30-38
3. von Neuman, J., H.R. Kent, H.R. Bellinson, and B.I. Hart, *"The Mean Square Successive Difference,"* Ann. Math. Stat., Vol. 12, pp. 221-331, 1942
4. Allan, D., Marc Weiss and J. Jespersen, *"A Frequency Domain View of Time-Domain Characterization of Clocks and Time and Frequency Distribution Systems."* 45th Annual Symposium on Frequency Control, May 1991
5. Rutman, J. and Walls, F., *"Characterization of Frequency Stability in Precision Frequency Sources."* Proc. IEEE, Special Issue on Time and Frequency, Vol. 79, pp. 952-960, July 91

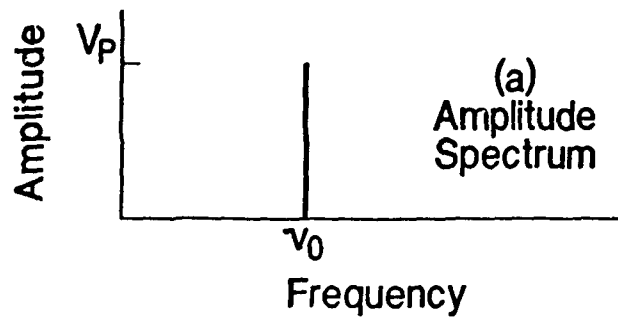


Figure 1a Signal Characterization of a Sinusoid in the Frequency Domain

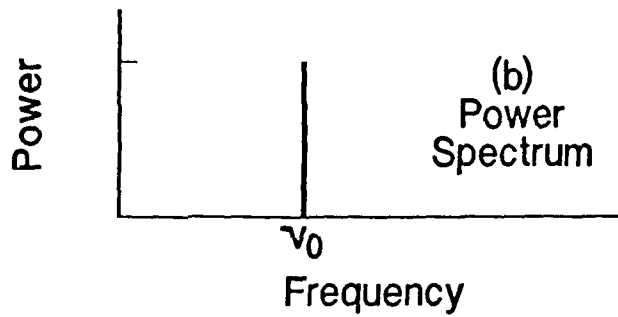


Figure 1b Power Characterization of a Sinusoid in the Frequency Domain

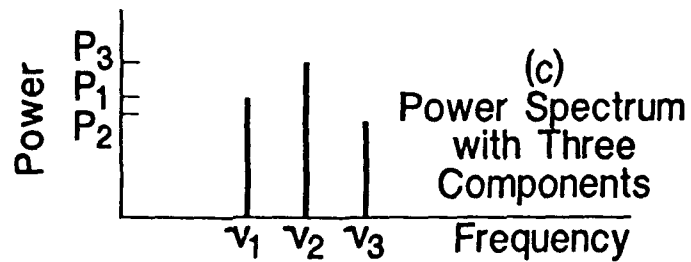


Figure 1c Power Characterization of Sinusoids at Three Different Frequencies in the Frequency Domain

Noise Power Spectrum

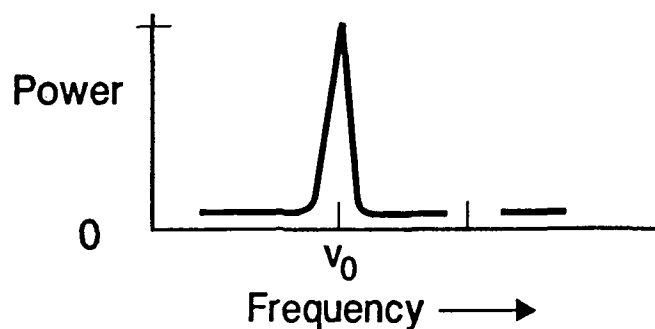


Figure 2 Power Spectrum of a Noisy Signal

Spectral Density vs. Fourier Frequency as Translated from Time Domain

Random Walk FM

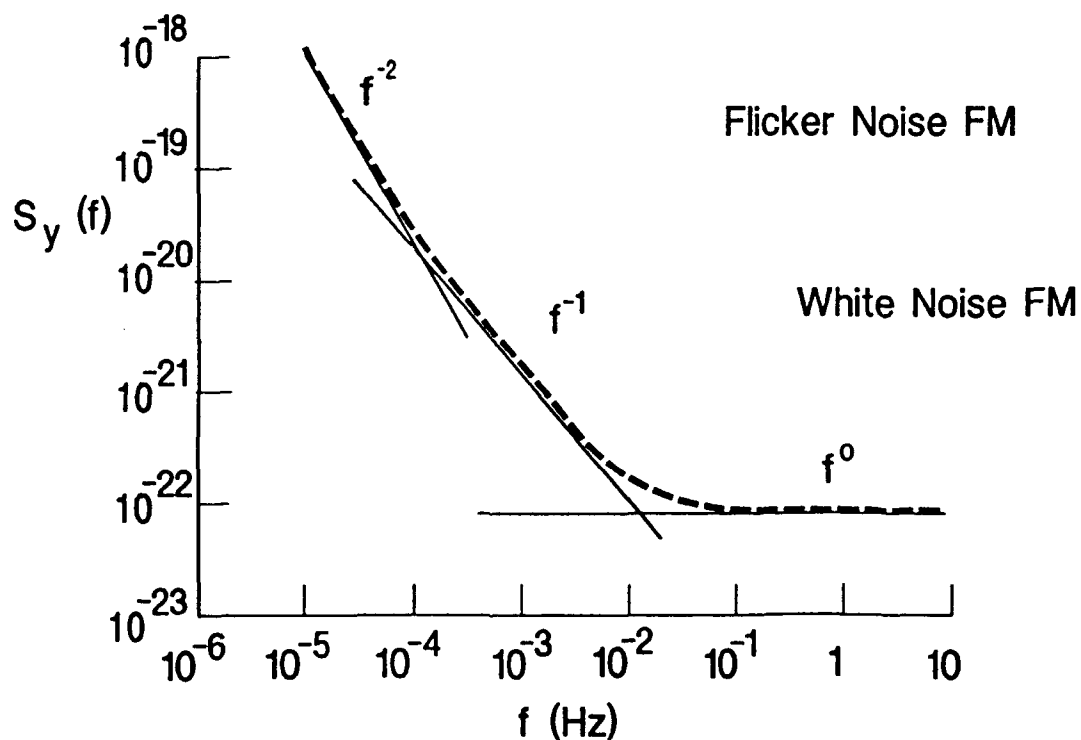


Figure 3 Power Spectral Density of a Typical Frequency Standard

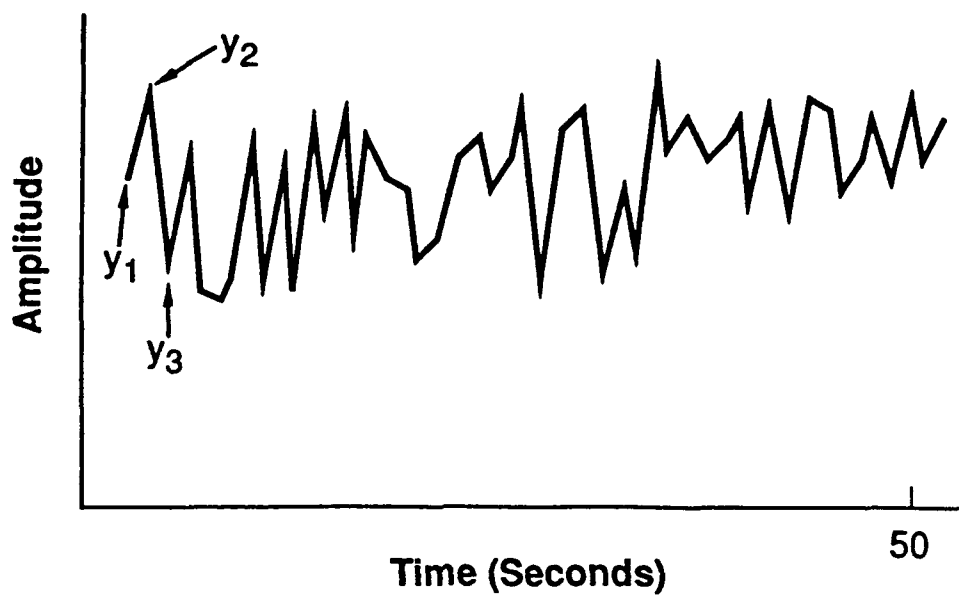


Figure 4a A Sampled Data Record

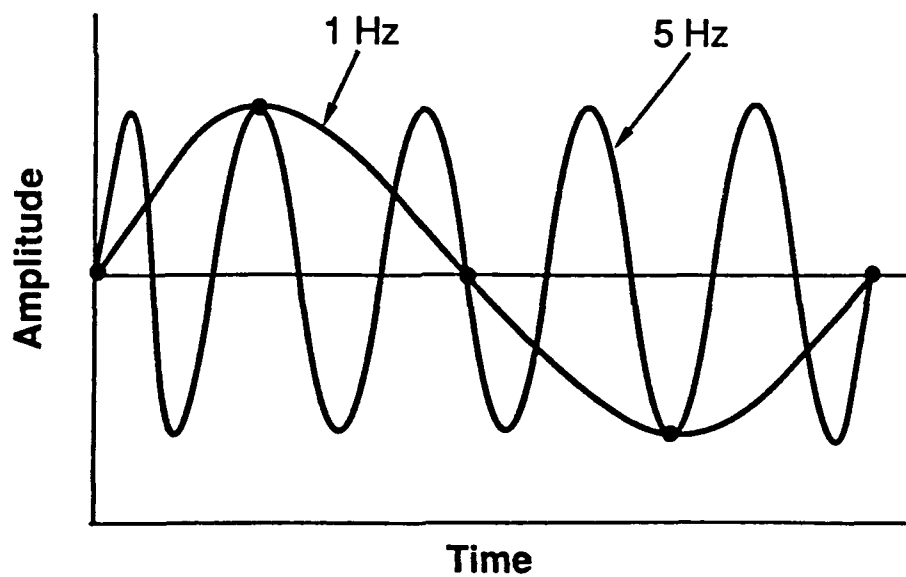


Figure 4b Confusing a 1 Hz with a 5 Hz Signal

High Pass Filter

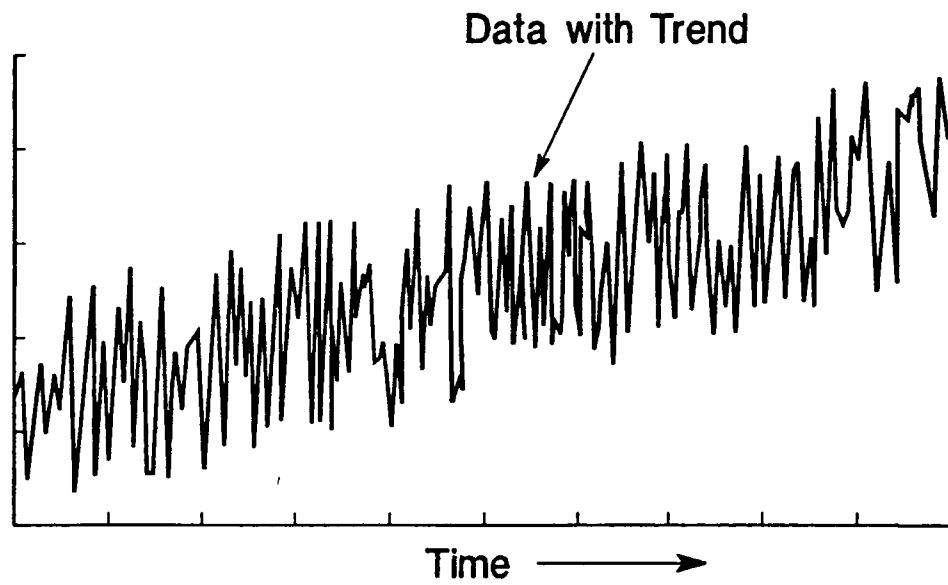


Figure 5a

Data with a Trend

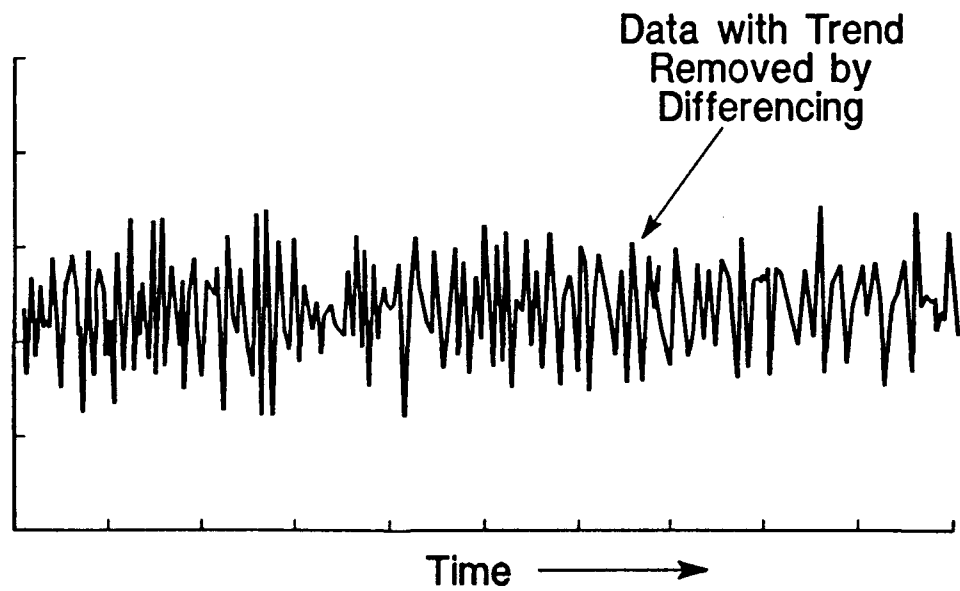


Figure 5b

Data with Trend Removed

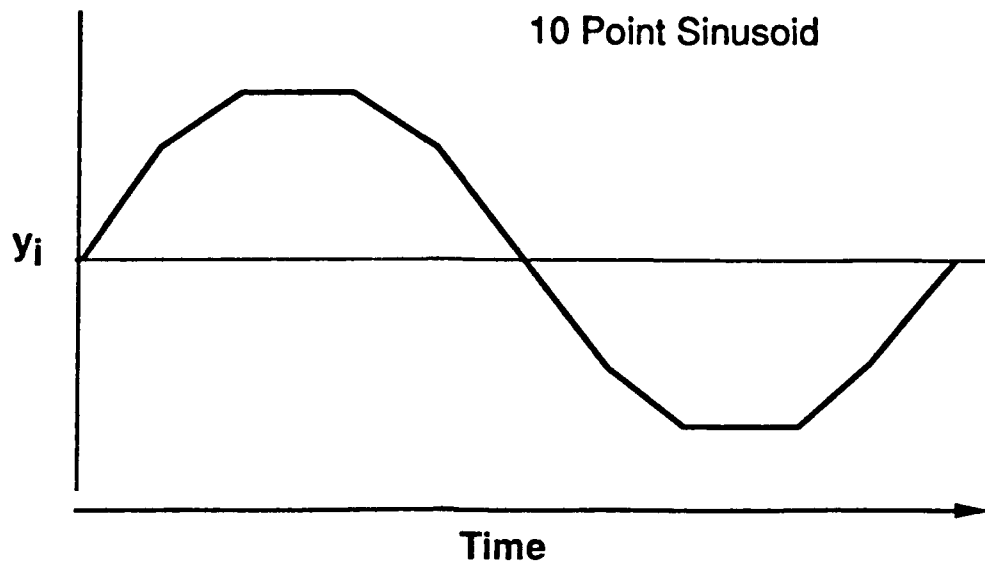


Figure 6a Original Signal

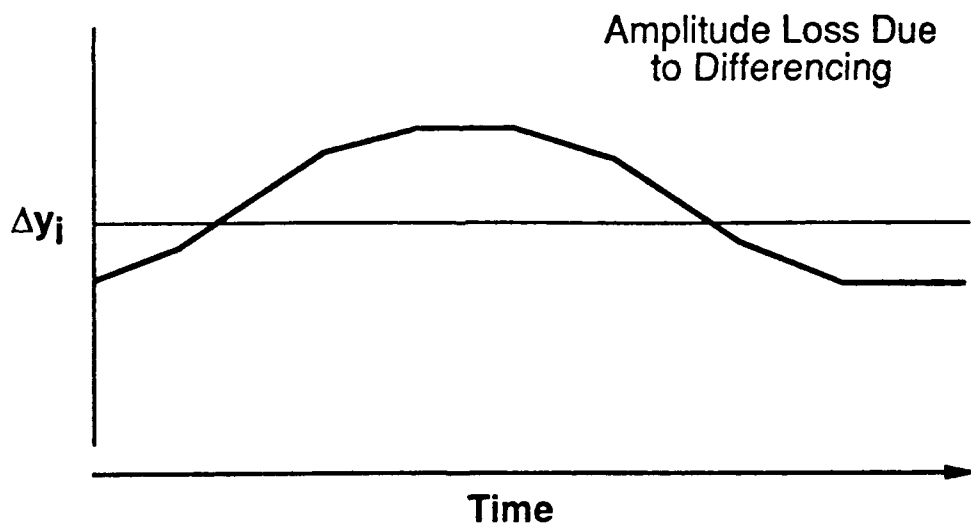


Figure 6b Signal after Differencing Data

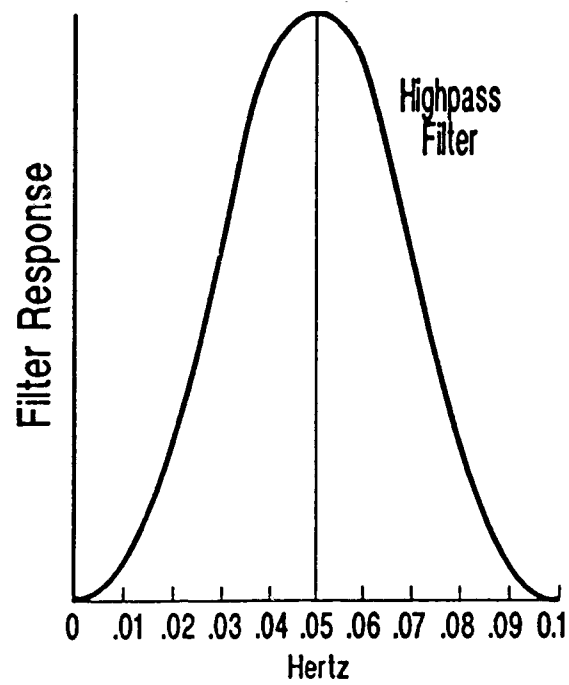


Figure 7a High-Pass Filter due to Differencing Data

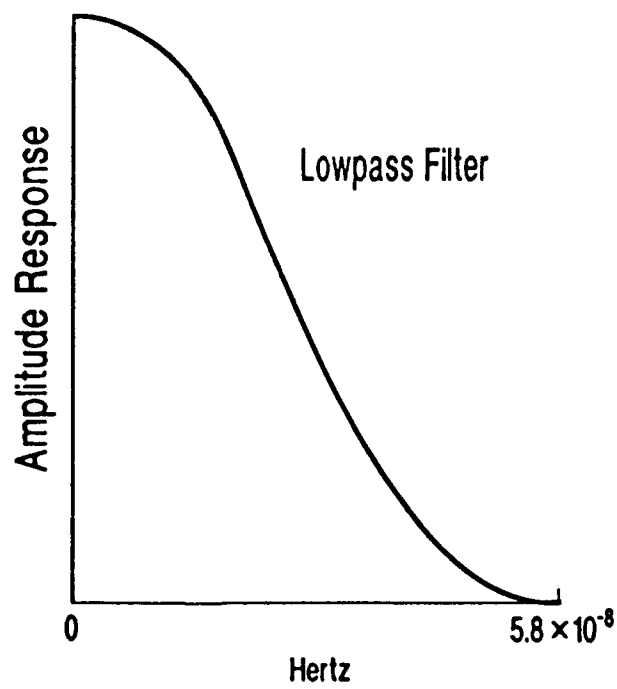


Figure 7b Low-Pass Filter due to Averaging Data

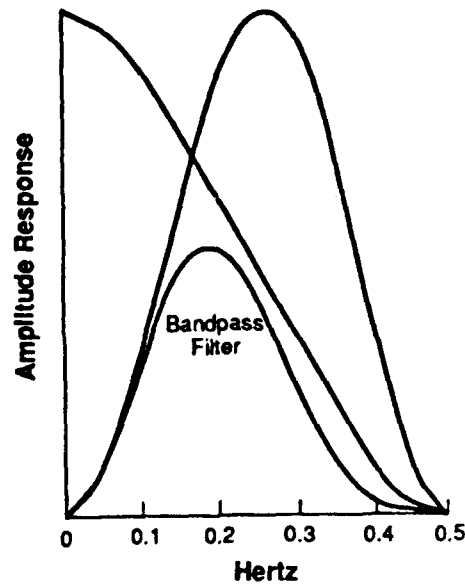


Figure 8a Band-Pass Filter due to Averaging and Differencing Data

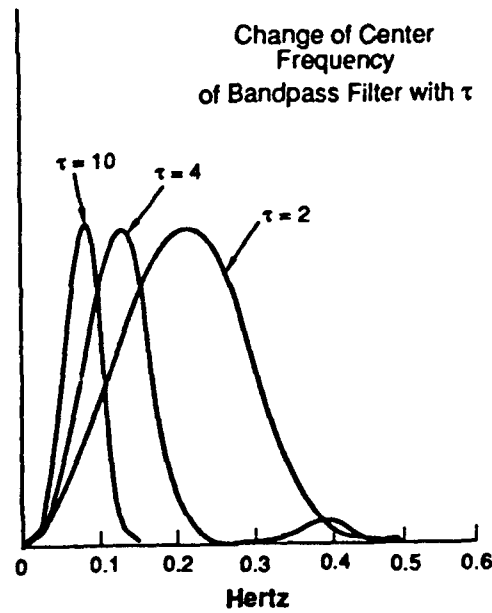


Figure 8b Change in Filter Characteristics due to Changing Averaging Time

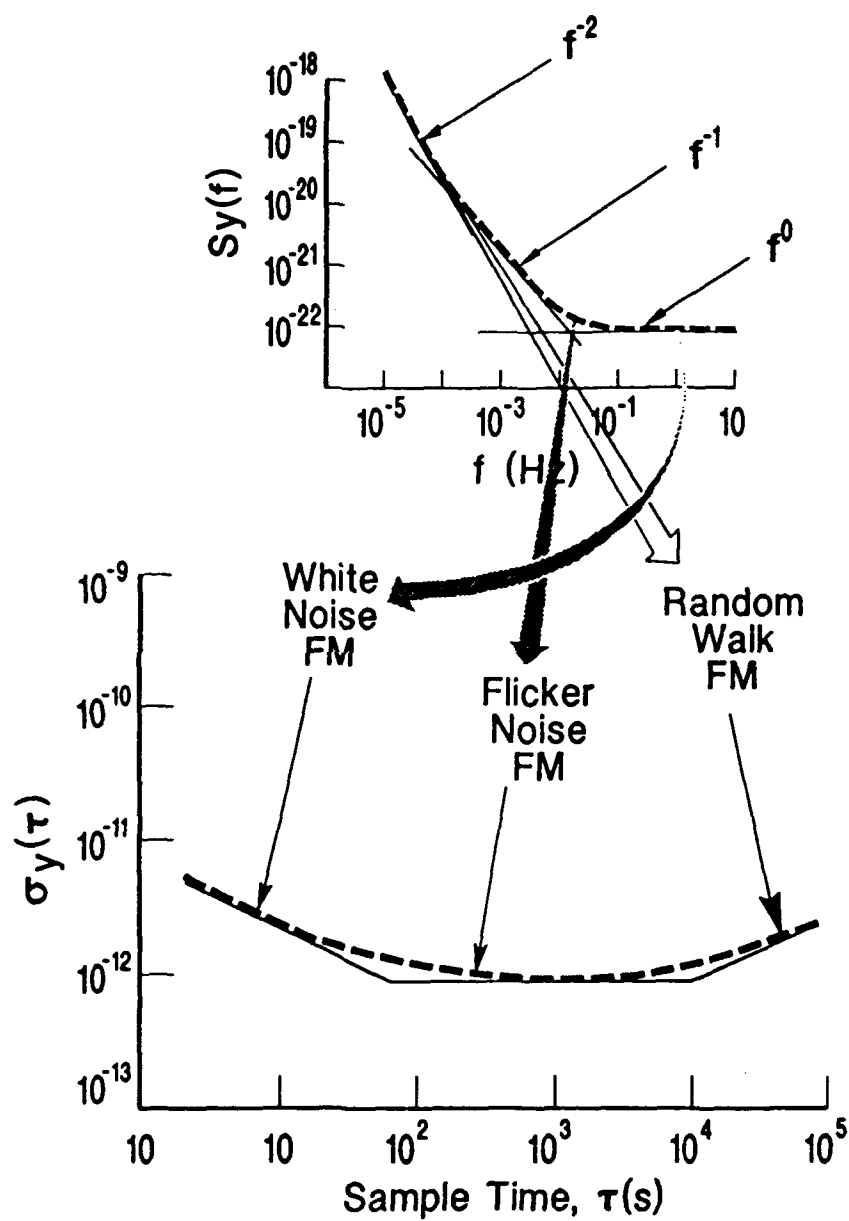


Figure 9 How Power Spectral Density and Pair Variance are Related

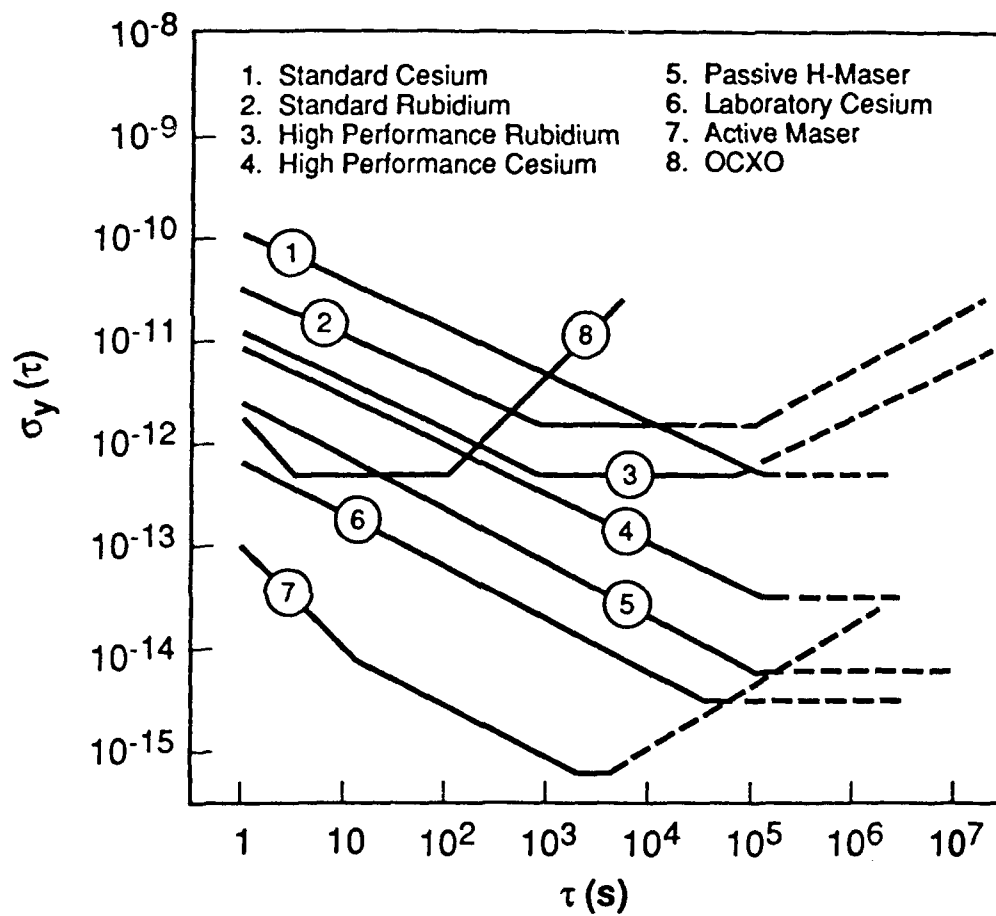


Figure 10 Characteristics of Typical Frequency Standards

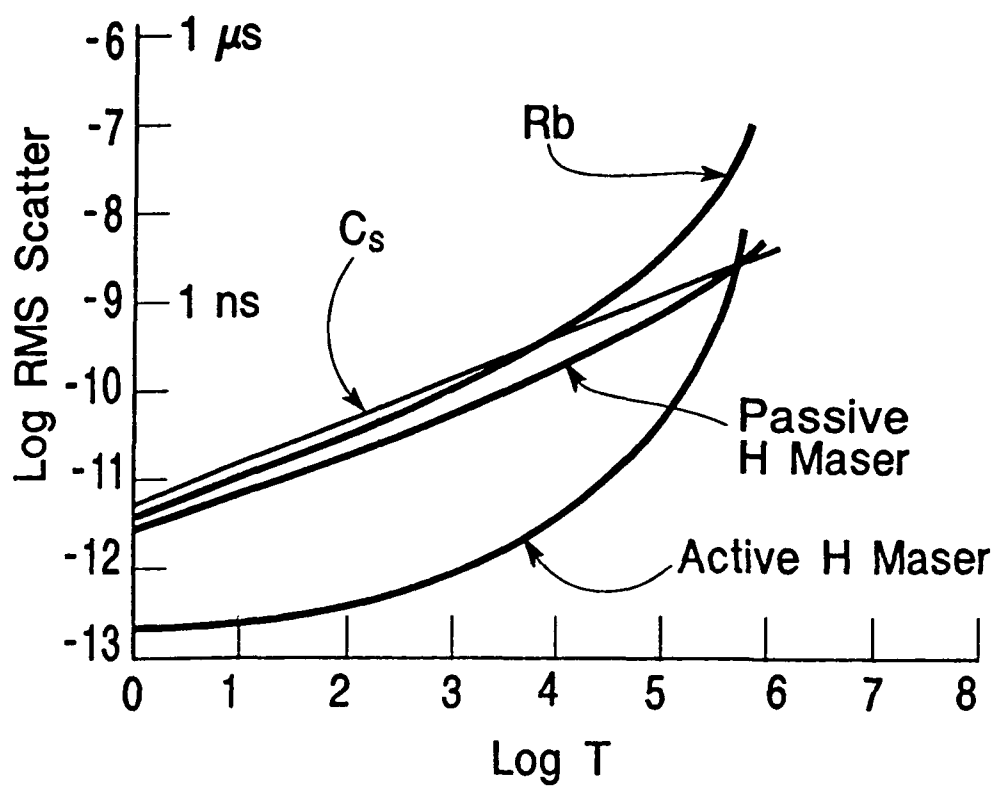


Figure 11 Predicting Clock Performance

NOISE MODELS FOR TIME AND FREQUENCY: A TUTORIAL

James A. Barnes
Austron, Inc

Abstract

The simple mean, the standard deviation and its square, the variance, are well known statistical measures of a set of data points. It seems natural, therefore, to apply these measures to signals generated by frequency standards. However, this approach quickly reveals some significant problems.

Normally, when we compute the mean and standard deviation of some process, we assume that including more data points in the computation brings us ever closer to the true mean and standard deviation of the process. Unfortunately this is not true, in general, for data obtained from signals generated by frequency standards. Here, the situation usually deteriorates as more data points are included.

This tutorial describes, largely in heuristic terms, modified statistical measures which do not "blow up" for the kinds of noise processes normally encountered in signals generated by frequency standards.

Introduction

At some level, random variations in measurements are universal. These variations are often referred to as "noise". There are many possible types of noise that become more or less important in various systems and configurations. Typically, two or three noise types are sufficient for a given situation. Some typical noises are: "White" noise, random walk noise, flicker noise, and "popcorn" noise to mention a few. One often sees systematic variations (like 60 Hz sidebands or a nearly constant frequency drift) included in the "noise" models. These deterministic signals are normally unwanted perturbations and it is reasonable to group them with the true, non-deterministic noises. For computer simulation studies one needs to be able to generate these noises (including some systematic effects) efficiently and reliably. This paper reviews some of the more common noise types, their simulation and some consequences.

White Noise

White light is light with all colors present in nearly equal amounts. Borrowing this idea of white light, one speaks of "white noise" when the Power Spectral Density (PSD) of the noise is nearly level over the range of Fourier frequencies of interest to the current work. This is an approximation for a limited bandwidth since, if white noise covered ALL frequencies, then the total power in the PSD would be infinite – clearly a non-physical result. In the audio range, the hiss of an air-leak is fairly white. In a digital system, white noise is a time series of random, non-correlated numbers.

Typically, the white noises are thought to be normally (or Gaussianly) distributed; that is, it is distributed according to the normal, bell-shaped curve. Most computers have algorithms to generate a sequence of nearly uncorrelated (i.e., white) pseudo-random numbers which are uniformly distributed over the interval [0,1]. These numbers can be used to generate a nearly Gaussian distribution, with zero mean, and unit variance by using the following algorithm (in BASIC):

```
200 p = 0 : FOR j = 1 TO 6 : p = p + RND - RND : NEXT j
```

where each call to RND returns the next pseudo-random number in the interval [0,1]. Upon completion of Line 200, the p-values have the desired properties. The numbers, RND, are said to be pseudo-random because the sequence of numbers will repeat exactly given the same "seed" number to begin the sequence. While they have the same sequence for a given seed, the auto-covariance is very nearly a delta function (i.e., a sharp spike with zeros around it) at zero lag-time. Also, for different seed numbers, the cross-correlations between time series are very nearly zero.

The General Noise Model

Figure 1 depicts a fairly general noise model capable of explaining many of the observed features of real clock data. This is a good place to emphasize that "good" models are simple models! One must avoid the temptation of over-modeling: That is, one should at most consider two to four model elements out of the many possible models. Figure 1 is an extreme case of over-kill. Of course, the purpose of Fig. 1 was to show many possibilities to consider. The book on time series analysis by Box and Jenkins [2] recommends strongly the use of "parsimonious" models. In general, if one gets only a small improvement by adding the N-plus-first element to the model, then one should NOT add the new element. One should note that an N-th order model can be fit to a time series of N values exactly – NO error! However, the forecast of the N-plus-first data point can be many, many orders of magnitude in error relative to the actual value when it becomes available. This is a case of working harder, not smarter. DON'T OVER-KILL.

Power-Law Noises

Fractals

By definition, if X_n is a white noise, then the power spectral density (PSD) of X_n is a constant, $2\sigma_a^2$, where σ_a^2 is the variance of the input shocks, a_n . In this case, the model is especially simple:

$$X_n = K a_n, \text{ for } n = 1, 2, 3, \dots, N \quad \text{and} \quad a_n \sim N(0, \sigma_a^2)$$

where K is a constant gain. The relation on the right side of the line above can be read as: "The a_n are normally distributed with zero mean and variance σ_a^2 (which is unity in our present discussions)". Correspondingly, the outputs, X_n , are distributed according to :

$$X_n \sim N(0, K^2)$$

since $\sigma_a^2 = 1$. Since the a_n 's are random uncorrelated (i.e., white), the PSD of X_n is $2K^2$ which can be written in the form $S_x(f) = 2K^2 f^0$ where f is the Fourier frequency and f^0 is a power-law with exponent zero.

The running sum (i.e., the finite integral) of a white noise has a power-law PSD which varies as f^{-2} . For the discrete case, this noise is known as the random walk. The random walk can be computed from the following relation:

$$X_n = X_{n-1} + a_n \quad \text{for } n = 1, 2, 3, \dots N.$$

For a continuous process, it is known as a Brownian motion and the sum would be replaced by an integral.

A second summation on X_n of the above summation would cause the PSD to vary as f^{-4} . In general, each summation lowers the exponent by another -2. Correspondingly, a first difference of the data, $\Delta X_n = X_n - X_{n-1}$, increases the power-law exponent of the PSD by +2.

In general any noise whose PSD varies as a power-law with even, integral exponent can be simulated by the appropriate differencing or summing of a random (white) time series.

The integration operator used in Laplace transforms is $1/s$ where $s = i\omega$. This opens the door to the construction of power-law PSD's with any exponent. For example, consider the Laplace operator $1/\sqrt{s} = 1/\sqrt{i\omega}$, which lies between white noise and random walk noise. This noise has a special name: "Flicker Noise". Its PSD varies as f^{-1} . The total power in flicker noise diverges at both the high end (the ultraviolet "catastrophe") and at the low end (the infrared "catastrophe") of the spectrum. Of course, this is not physically reasonable. Indeed, if I purchased a device that put out infinite power I would immediately return it to the dealer since it would fail to meet specifications! In reality, we observe signals which are bounded on both the high frequency end and the low frequency end. The "catastrophes" are not real.

Technically, a random walk is a discrete process which is the running sum of random, uncorrelated shocks. A Brownian motion is the continuous equivalent of the random walk involving the integral of a continuous random noise. In real situations one must be aware of finite pass-bands of real equipment and recognize that the data are always finite as well as their PSD's [3]. Occasionally, one encounters a criticism of the power-law noise elements because of the infinities at either zero or infinite Fourier frequencies. These arguments are specious since one can never observe these infinities.

In a very real sense the oscillator performance is pre-conditioned to meet the technical specifications given by the manufacturer at the time of sale.

Combining the ideas of fractional integration and Brownian motion, Mandelbrot coined the name "Fractals" for fractional Brownian motions.

Self-Similar Processes

There is another concept which is realizing significant interest - self-similar processes. Imagine a large box containing thousands of sheets of paper which can be stretched in both directions (length and width). On each sheet of paper is a plot of a realization of the same specific fractal process with each sample of the same length. Each sample was produced by exactly the same prescription -

only the “seed numbers” for the random noise generators were different were different from sample to sample. Imagine that we take samples at random from the box and with great care analyse the data and discover that everything is consistent with some specific fractal model.

Now let’s take samples from the first box, stretch them along the time axis until they are twice as long as they were originally, and cut it back to its original length. We will need to stretch the Y-axis also but the amount of stretching is yet to be determined. A large number of the modified plots will be accumulated (after the Y-axis adjustments have been made). As before, we analyse this data and get reasonable concurrence with the previous analysis of the first box. If the statistics of the two boxes are essentially the same, then the original process is said to be “Self-Similar”.

Let’s assume that the original noise was a random walk and one knows that a random walk grows as the square root of the data length. Thus, if one stretches the time axis a factor of, say, T , then the vertical axis, V , should be scaled by the square root of T . The ratio, R , of V/T should be constant for any T . In the above example $R = V/T = 0.7071\dots$, in accord with the random walk assumption. All of the power-law noises are self-similar as well as being fractals. As one might expect, the ratio, R , is different for each power-law. For white noise, the vertical scaling factor is a constant, since the vertical excursions don’t change with T . For flicker noise, f^{-1} , the vertical scaling factor is proportional to $\sqrt{\ln 2} = 0.8326$. For an f^{-3} process, the vertical and horizontal scale factors are identical (i.e., $R = 1$). Thus, putting an f^{-3} process under a microscope would leave the statistics unchanged regardless of the microscope power. Finally, the double sum of random numbers (i.e., the integral of a random walk) has an f^{-4} noise and an R -ratio given by $\sqrt{3/2} = 1.2247$. Scientists have spent a great deal of time looking over old data on such things as the length of the day, sunspots, and global warming, to name three. If one of these is described well as a self-similar process, say, an f^{-3} process, more data will not improve the statistics! For an f^{-4} process, the variances grow with increased quantities of data. For f^0 , f^{-1} , and f^{-2} processes, more data improves the confidence intervals. It is fascinating to think that some of these processes might be unknowable in a statistical sense. (Godel’s Theorem?)

The primary model for sinusoidal signals from quality oscillators or other signal processing equipment is as follows: $V(t) = V_0 \sin(\Phi(t))$ where V_0 is the nominal, constant amplitude of the sinusoid and $\Phi(t)$ is the total phase, given by:

$$\Phi(t) = 2\pi\nu_0 t + \phi(t)$$

where ν_0 is the nominal frequency in Hz and $\phi(t)$ (in radians) contains all of the random components in time. For consistency, we define $x(t)$ and $y(t)$ by the relations:

$$\begin{aligned} x(t) &= \frac{1}{2\pi\nu_0} \phi \\ y(t) &= \frac{1}{2\pi\nu_0} \frac{d\phi}{dt} \end{aligned}$$

and by definition,

$$y(t) = \frac{dx}{dt}$$

In words, $x(t)$ is the relative time error (measured in seconds) of a clock driven by the signal, $V(t)$, to some hypothetical standard. $y(t)$ is the relative frequency error (usually expressed as a fraction - e.g., 3.5×10^{-12}). Another useful equation is:

$$y(t) = \frac{\nu(t) - \nu_0}{\nu_0}$$

One can imagine tuning an oscillator's frequency with some random (or non-random) signal. Also, one can imagine changing the phase of a sinusoidal signal in a random (or non-random) way. One can use the power-law noises to modulate either the phase or frequency or both. From this one can construct the five most common power-law noises:

| NAME | $S_x(f)$ | $S_y(f)$ |
|------------|------------------------------------|-----------------|
| WHITE PM | $\frac{1}{(2\pi)^2} h_2 f^0$ | $h_2 f^2$ |
| FLICKER PM | $\frac{1}{(2\pi)^2} h_1 f^{-1}$ | $h_1 f^1$ |
| WHITE FM | $\frac{1}{(2\pi)^2} h_0 f^{-2}$ | $h_0 f^0$ |
| FLICKER FM | $\frac{1}{(2\pi)^2} h_{-1} f^{-3}$ | $h_{-1} f^{-1}$ |
| RND WLK FM | $\frac{1}{(2\pi)^2} h_{-2} f^{-4}$ | $h_{-2} f^{-2}$ |

Where use has been made of the relation:

$$S_y(f) = (2\pi f)^2 S_x(f) \text{ for } 0 < f < 1/2$$

Most clock and oscillator models incorporate one or two of these power-law noises as elements in their models. While non-integer values of the exponent for power-law noises are possible, in practice they are seldom used.

Other Model Elements

In addition to the power-law elements one often finds parts of the clock/oscillator performance similar to a quadratic polynomial:

$$X(nt_0) = X_o + Y_o n + D_r \frac{n^2}{2} + \epsilon_n \quad \text{for } n = 1, 2, 3, \dots, N$$

where $X(nt_0)$ is the time error for each value of n regressed on the initial time error, X_o , the relative frequency error, Y_o , and the linear frequency drift, D_r . For the regression to be valid, the residual errors given by:

$$\epsilon_n = \frac{\phi(nt_0)}{2\pi\nu_0}$$

must be random, uncorrelated (i.e., white). The only noise that satisfies this condition is just white phase modulation (PM). Of course one can perform the mechanism of regression analysis to the data any way and simply choose to ignore the "goodness of fit". With other power-law noises the problems get worse. If one includes random walk FM in the model, it becomes very difficult indeed to get any reasonable confidence in the analysis.

Of course, there are other noises in real problems than just the power-law noises. One such noise is "popcorn" noise, named after the noise generated by corn popping. It is also referred to as a random telegraph signal. Typically the noise makes discrete jumps in value at an audio rate. The popcorn noise source/s are not known. It is possible for a popcorn noise to have a PSD which is approximately flicker noise ($1/f$) over a substantial Fourier frequency range.

Sporadic noises or noise bursts also can exist, and may not be altogether different from popcorn noise.

There are also noises which suffer from heteroschedasticity. That is a noise whose variance is not constant. For example, the variance of a received VLF radio signal has different noise levels corresponding to daylight on the path or night time on the path. This is a family of noises which have been ignored in the past. (I expect the lack of interest in these noises is due for change in the not-too-distant future.

Many noises have their explanation in environmental effects: temperature, humidity, vibration, EMI, power supply problems and many more.

Simulation

There are three primary values in simulating data:

1. It forces the experimenter to state his model assumptions explicitly;
2. It provides a check on the adequacy of the model; and
3. It provides an independent means of verifying the analysis routines.

There is also a significant pitfall that could be disastrous: The confusion of simulated data samples with real data.

References

1. Barnes, J.A., (1991), *"The Analysis of Frequency and Time Data"*, Austron, Inc.
2. Box, G.E.P. and G.M. Jenkins, (1970), *"Time Series Analysis, Forecasting and Control"*, Holden-Day San Francisco.
3. Slepian, David (1976), *"On Bandwidth"*, Proc.IEEE, 64(3),292-300.

GPS Time Determination and Dissemination

–A Tutorial–

S. R. Stein
Timing Solutions Corporation

T. R. Bartholomew
The Analytic Sciences Corporation (TASC)

Lieutenant C. Fox
United States Air Force

G. A. Gifford
Naval Research Laboratory

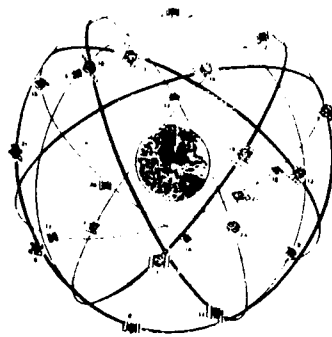
GPS SYSTEM CONCEPT

The Global Positioning System (GPS) is a U.S. Department of Defense (DoD) satellite-based navigation system that provides precise navigation, time recovery, and frequency control capabilities to users located anywhere in the world. The current system plan calls for the deployment of 21 satellites, three of which are active spares. The operational satellites, which are designated as Block 2, are being launched to six planes at fifty-five degrees inclination. These satellites radiate time-coded signals and information giving the time of an on-board atomic clock and the position of the satellite. All users are passive. They receive these time-coded signals from the satellites and use them to compute 3-D position and time of their local clock. Users make no transmissions to the satellite and need not communicate with each other in order to access the system.

PERFORMANCE REQUIREMENTS

The system performance requirements are set forth in the GPS Program Management Directive (PMD). The system is required to deliver navigation accuracy of sixteen meters Spherical Error Probable (SEP). This means that fifty percent of the time the navigation error will be less than sixteen meters. The system's internal time scale is called GPS Time. It must be maintained within one microsecond of Universal Coordinated Time (UTC), that is UTC(USNO) disregarding the integer number of seconds offset from UTC caused by the irregular leap seconds used to keep UTC coordinated with UT1 or Earth's time. Finally, GPS is required to provide the user with correction information in the navigation message so that the user may compute the offset from GPS to UTC(USNO) accurate to within 110 ns (1 sigma).

GPS SYSTEM CONCEPT



- 21 satellites, 3 active spares in 6 10,900 N. mile (12 hr) orbits
- Block I — 2 planes, 63° inclination
- Block II — 6 planes, 55° inclination, 2 Cs and 2 Rb Clocks
- Satellites radiate time coded signals
- Passive users can determine 3D position and time
- World-wide coverage

PERFORMANCE REQUIREMENTS

- 16 meter spherical error probable (SEP) navigation accuracy
- maintain GPS time within 1 μ s of UTC(USNO) except leap seconds
- provide GPS - UTC(USNO) offset accurate to 110 ns

GPS CONFIGURATION DECEMBER 1991.

As of this writing, December 1991, there are sixteen satellites in the GPS constellation. These satellites are a mixture of Block 1, Block 2, and Block 2 survivable satellites and broadcast their time-coded signals as derived from either cesium or rubidium atomic clocks. The spacecraft cesium clocks and the ground-based monitor station cesium clocks are now used to compute GPS Time using an algorithm known as the Composite Clock.

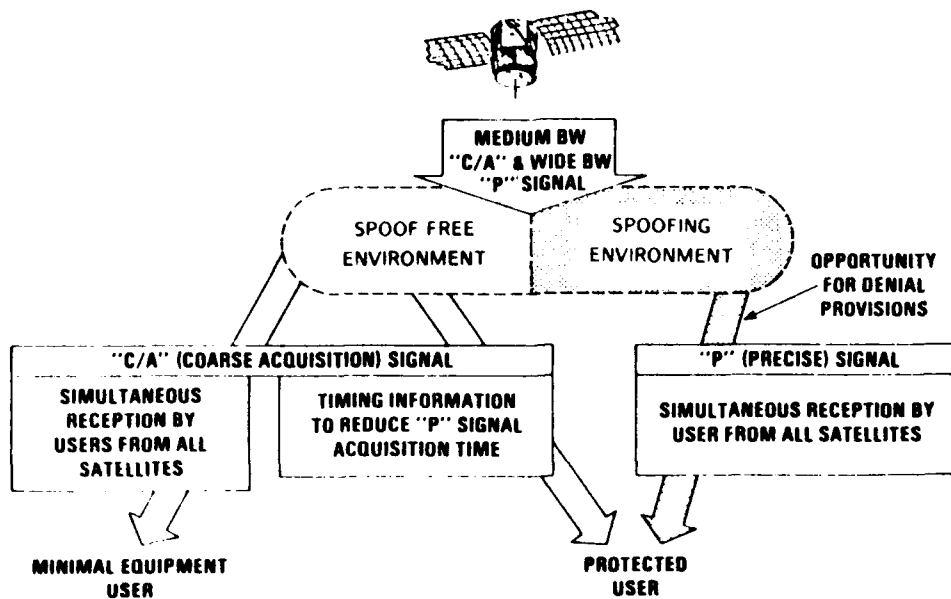
COARSE AND PRECISE SIGNAL

GPS satellites transmit two independent timing signals. The first is a C/A code or Course Acquisition signal utilized by all Standard Positioning Service (SPS) users. The second is a P code or Precise signal that may generally be accessed only by users of the Precise Positioning Service (PPS). SPS C/A code users can navigate to approximately 100 meter (2 DRMS) accuracies and recover time to accuracies of about 250-300 ns (1 sigma). Users of the P code utilize the C/A code to reduce the time required to access the P code. In addition to providing the full accuracy capability of the GPS system, the PPS provides authorized users with protection from spoofing signals. With its higher process gain, the P code also provides additional immunity from CW-type jammers. However, the C/A code actually provides more immunity from broad-band jammers due to the higher power in this signal at the user antenna.

GPS CONFIGURATION DECEMBER 1991

| PLANE | SVN | PRN | BLOCK | CLOCK TYPE | COMPOSITE CLOCK |
|-------|-----|-----|-------|------------|-----------------|
| A1 | 10 | 12 | I | Cs | No |
| A3 | 3 | 6 | I | Rb | No |
| A4 | 19 | 19 | IIN | Cs | Yes |
| B2 | 20 | 20 | IIN | Cs | Yes |
| B3 | 13 | 2 | IIN | Cs | Yes |
| C1 | 9 | 13 | I | Cs | Yes |
| C3 | 8 | 11 | I | Cs | Yes |
| C4 | 11 | 3 | I | Rb | No |
| D1 | 24 | 24 | IIS | Cs | Yes |
| D2 | 15 | 15 | IIN | Cs | Yes |
| D3 | 17 | 17 | IIN | Cs | Yes |
| E1 | 14 | 14 | IIN | Cs | Yes |
| E2 | 21 | 21 | IIN | Cs | Yes |
| E3 | 16 | 16 | IIN | Cs | Yes |
| E4 | 23 | 23 | IIS | Cs | Yes |
| F3 | 18 | 18 | IIN | Cs | Yes |

COARSE AND PRECISE SIGNAL



GPS SIGNALS

GPS uses two L-band microwave frequencies for transmitting the navigation signals. The L1 carrier at 1575 MHz is the primary navigation signal and carries both the C/A and P codes. The L2 carrier at 1228 MHz is a secondary navigation signal. Although it may be switched between the C/A or P code, it can never carry both at the same time and in practice will always carry the P code. The purpose of the L2 carrier is to permit PPS users (and codeless SPS users) to obtain higher accuracy ionospheric delay calibration than C/A users. The L3 carrier at 1381 MHz is not offered to the navigation user but rather carries signals utilized for global burst detection. Finally, command traffic from the ground to the GPS satellites is carried on a standard S-band SGLS datalink.

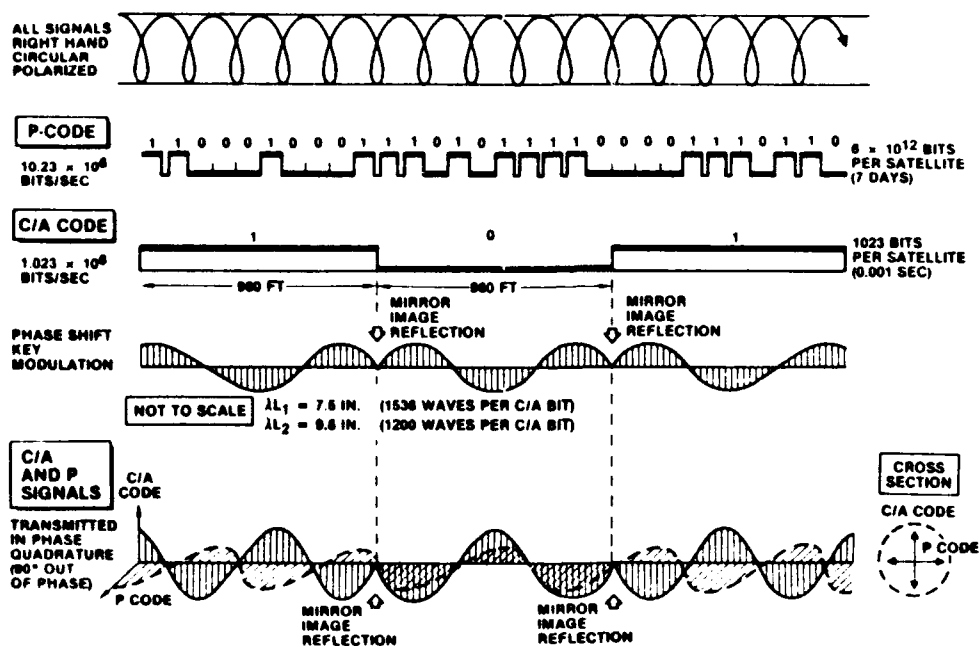
SIGNAL MODULATION TECHNIQUES

The C/A code has a data rate of 1.023 megabits per second and a code length of 1023 bits. The C/A codes are Gold codes and each satellite uses an independent code which is approximately orthogonal to all the others. The data rate of the P code is ten times that of the C/A code or 10.23 megabits per second. There is a single P code and the satellites are distinguished by offsetting the origin of the code in each satellite. The total code length is 267 days and the offset from one satellite to the next is 7 days. All codes are phase-shift-keyed onto the carrier which is right-hand-circularly polarized and the C/A and P signals are transmitted in phase quadrature from one another.

GPS SIGNALS

- **L₁ — 1575.42 MHz**
 - Primary navigation signal
 - C/A and P Codes and navigation data
- **L₂ — 1227.6 MHz**
 - Second frequency provides higher accuracy ionospheric delay calibration
 - P Code and navigation data
- **L₃ — 1381.05 MHz**
 - Global burst detector
- **Command channel — S-Band**

SIGNAL MODULATION TECHNIQUES



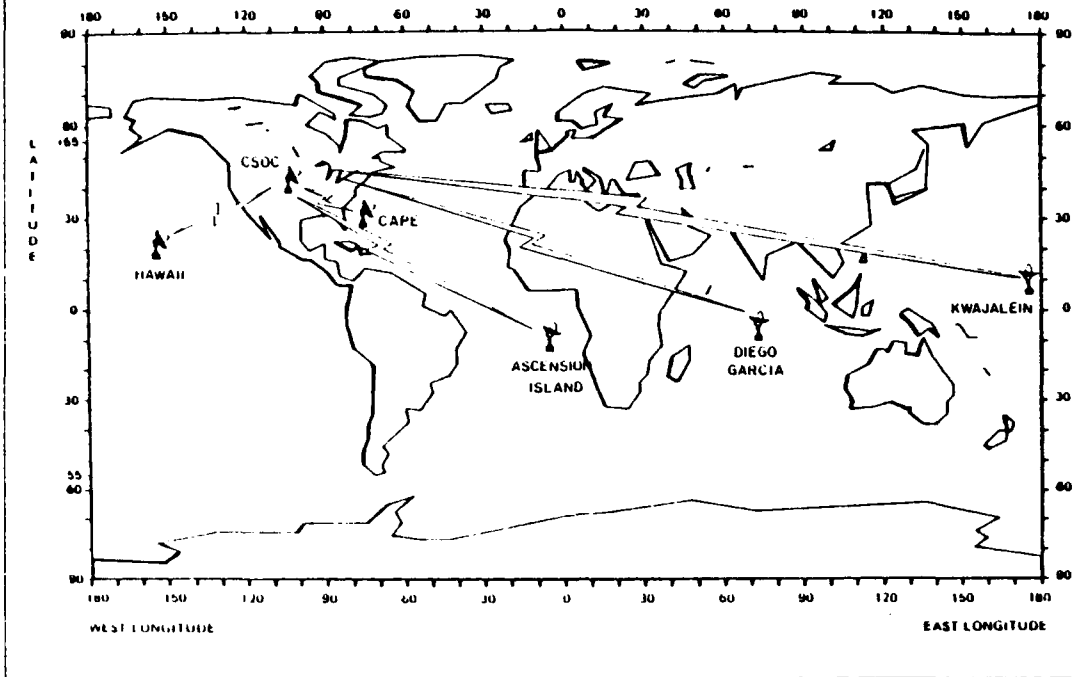
OPERATIONAL CONTROL SEGMENT

GPS is operated from a Master Control Station (MCS) located at the Combined Space Operation Center (CSOC) at Falcon Air Force Station in Colorado Springs in conjunction with globally distributed monitor stations and ground antennas. The MCS and the monitor stations are collectively known as the Control Segment. A GPS user navigates or receives time by performing ranging measurements to the known positions of the satellites.

PURPOSE OF THE CONTROL SEGMENT

The purpose of the Control Segment is to enable the satellites to accurately tell users where the satellites are and when each satellite transmitted its signal. To accomplish this, the monitor stations continuously make pseudo-range measurements from the satellites whenever they are visible. These data are sent to the MCS which computes the satellite ephemerides and both the satellite and monitor station clock parameters. The ephemeris and clock parameters for each satellite are uploaded daily from the MCS, and finally these data are made available to the user in the navigation message transmitted by the satellite on the L-band signals.

OPERATIONAL CONTROL SEGMENT



PURPOSE OF THE CONTROL SEGMENT

- Monitor stations (MS) continuously make pseudorange measurements from the satellites whenever they are visible, similar to high quality user receiver
- Data sent to master control station (MCS)
- MCS computes satellite ephemeris and both SV and MS clock parameters
- Ephemeris and clock parameters uploaded by ground antenna (GA)
- Data stream from the SV to the user contains updated orbital elements and clock parameters for that satellite
- PURPOSE: Tell user where the SV is and when its signal was transmitted

ROLE OF THE MASTER CONTROL STATION

The MCS estimates the orbit and the time of the clock for each satellite. The MCS controls all the monitor stations which themselves are unmanned. The data transmitted by the monitor stations are collected and processed in 15-minute batches. First, the data are corrected for bad measurements and then they are smoothed. Finally, they are utilized to update the current estimates of the orbit, the satellite clock times, and the monitor station clock times using a Kalman filter. The current state estimate is used to calculate predictions for both the satellite orbit and clock time, and the information is uploaded to the satellites from the MCS.

COMPOSITE CLOCK

Operations Release 5.2 of the Operational Control System (OCS) changed the method for generating the GPS time which is transmitted by the satellites. GPS time is now computed as a weighted mean of both satellite and monitor station clocks. This computation is designated the Composite Clock and should be more stable than any of the individual clocks that make up the ensemble. Some of the more poorly performing satellite clocks, such as the rubidium clocks and misbehaving cesium clocks, are excluded from the Composite Clock by placing them in separate partitions which are not used in the Composite Clock calculation. Current performance data indicate that the Composite Clock exhibits approximately the same stability as a single cesium clock. This performance is not as good as expected. However, the OCS is still obtaining significant advantages from the use of the Composite Clock since it has greatly reduced the number of disturbances in GPS time which previously resulted from disturbances in the Master Monitor Station clock or changes in the designation of Master Monitor Station.

ROLE OF THE MASTER CONTROL STATION

- **BASIC FUNCTION:** Estimate SV ephemeris and clock states
- **FUNCTIONS**
 - Control MS and receive data
 - Monitor system status
 - Correct and smooth MS pseudorange data
 - Generate estimates of SV ephemeris and SV and MS clock time, frequency, and frequency aging using Kalman filter
 - Generate predictions of clock and ephemeris
 - Format into navigation message and upload blocks
 - Control ground antenna

COMPOSITE CLOCK

- GPS time is now a weighted mean of a subset of the clocks in the system including all the MS clocks and some of the SV clocks
- The ensemble mean should be more stable than any individual clock in the system
- Partition reconciliation is needed to ensure that GPS time is consistent across all partitions
- Some weighting is done in the partition reconciliation process
- Excluded clocks are in partitions with 0 weight
- Ensemble performance in 1991
 - Stability equal to a single Cs clock
 - Reduced number of disturbances attributable to MS clocks

ROLE OF THE MONITOR STATIONS

There are five monitor stations located at Hawaii, Ascension Island, Diego Garcia, Kwajalein, and Colorado Springs. They collect pseudorange data for the MCS. Monitor stations are unmanned. A control computer operates the monitor station under command from the MCS. Each monitor station is equipped with twelve-channel GPS receivers and redundant cesium beam frequency standards for maintaining the monitor station time. In addition to collecting pseudoranges, each monitor station collects the navigation messages which are used by the MCS to verify the accuracy of the upload data.

NAVY CLOCK ENSEMBLE

The clock for the Colorado Springs Monitor Station consists of an ensemble of Hewlett Packard HP5061 cesium clocks. A time scale computed from this ensemble is the weighted average of the times of all member clocks. Although conceptually similar to the Composite Clock used to compute GPS time, the Colorado Springs Monitor Station ensemble which was installed by the US Naval Observatory (USNO) and the Naval Research Laboratory (NRL) utilizes the KAS-1 algorithm which provides performance better than the best single clock in the ensemble. This system is now being used as the primary monitor station clock in Colorado Springs and its time is input to the GPS Kalman filter. Thus, this high performance time source will be available to evaluate the performance of other clocks in the GPS system.

ROLE OF THE MONITOR STATIONS

- 5 MS at Hawaii, Ascension, Diego Garcia, Kwajalein, and Colorado Springs
- Collect pseudorange data for the MCS
- Stations are unmanned — controlled entirely by MCS
- MS collect navigation messages for verification of uploads
- Key components:
 - GPS receiver (12 Channels)
 - Redundant Cs clocks
 - Control computer

NAVY CLOCK ENSEMBLE

- Located at the MCS in Colorado Springs
- Contains 6 HP Cs clocks
- Utilizes KAS-1 algorithm
- Fully automated
- Ensemble performance is better than the best member clock
- Now being used as the primary Colorado Springs MS clock

ROLE OF THE GROUND ANTENNA

Four ground antennas located at Ascension Island, Diego Garcia, Kwajalein, and Cape Canaveral may also be used to upload the navigation message data to the satellites. The ground antennas utilize ten-meter dish antennas at S-band to upload the navigation data when required. They may also be used to command various space vehicle functions such as station-keeping maneuvers. The accuracy of any uploaded data is verified through the monitor stations using the standard L-band downlink.

GPS ROLES OF OTHER ORGANIZATIONS

Several external organizations are involved in the operation and management of the Global Positioning System. The US Naval Observatory provides the calibration data for the offset between UTC(USNO) and GPS time. This data is used by the MCS to steer GPS time to Coordinated Universal Time (UTC) and to estimate the residual error in the steering process which is uploaded to the satellites. The USNO also maintains the hardware ensemble that is currently being used as the Colorado Springs Monitor Station clock. The Defense Mapping Agency (DMA) provides the MCS with polar orientation data needed to compute the real time estimates of the space vehicle (SV) orbits. The DMA receives GPS monitor station data from the MCS after smoothing and also collects its own data from a separate network of monitor stations located around the world. Using these data, DMA computes a precise ephemeris which is available after some delay. The precise ephemeris is used to evaluate the performance of the GPS real time Kalman filter and to provide corrections for various high precision positioning and time users.

ROLE OF THE GROUND ANTENNA

- BASIC FUNCTION: Transmit navigation message data to satellite
- 4 GA located at Ascension, Diego Garcia, Kwajalein, and Cape Canaveral
- Uplink compatible with standard SGLS (S-Band)
- 10 meter dish
- Upload verification by L-band downlink through MS
- Command SVs as required

GPS ROLES OF OTHER ORGANIZATIONS

- U. S. NAVAL OBSERVATORY (USNO)
 - Provides UTC(USNO) vs GPS calibration data
 - Maintains hardware ensemble at MCS
- Defense Mapping Agency (DMA)
 - Provides MCS with polar orientation data
 - Calculates precise ephemeris from 15 minute smoothed pseudorange measurements
- AFTAC
 - Receives daily global burst detector data

METHOD OF ESTABLISHING GPS AND SV TIME

The GPS time transmitted by the satellites is the true end product of the entire Global Positioning System. Together with the position of the satellite, it is used to compute the spatial coordinates and time of the users. For this reason it is very important to understand how GPS time is established and its relationship to satellite time. The time of the satellite clock is generated from the continuous output of the onboard atomic frequency standard and a counter which maintains the "Z" count of the SV. One Z count is approximately one and one-half seconds. Although the output of the onboard frequency standard is continuous, the MCS occasionally resets the Z count for convenience in handling the large numbers. GPS time is the coordinate time in the special relativity sense. It is computed using the composite clock algorithm as the weighted average of all the monitor station clocks and a subset of the satellite clocks. However, GPS time is not an independent time scale. Rather, it is steered to Coordinated Universal Time as maintained at the USNO. The steering corrections are implemented within the GPS system by changing the first derivative of the frequency of all the clocks within the system by an equal amount. This causes every clock in the system to experience an equal frequency aging. Since only the time difference between various clocks in the system may be observed through the monitor stations, the steering process has no effect whatsoever on internal operations. Its only effect is that GPS time ages with respect to any clock external to the GPS system. The aging corrections are generated by an automatic steering algorithm based upon a single smooth estimate of the GPS time offset from UTC(USNO) which is generated each day by the Naval Observatory. The algorithm either leaves the current frequency aging of GPS time unchanged or changes it by $\pm 2 \times 10^{-19} \text{ seconds/seconds}^2$. The algorithm operates without intervention by the MCS with the goal of maintaining GPS time within one microsecond of UTC(USNO). The UTC versus GPS offset data is also used by the MCS to compute the estimated offset of GPS from UTC and forecast that offset into the future. These data are uploaded to the satellites and broadcast in the navigation message so that each user may obtain more precise estimates of the offset of GPS time from UTC.

MCS COMPUTATIONS

The MCS estimates the satellite orbits and the times of the satellite clocks. The estimation is performed using a linearized and partitioned Kalman filter. The states actually estimated by this Kalman filter are the satellite ephemerides consisting of six orbital elements per satellite, the solar pressure constants of which there are two per satellite, the bias or time of the frequency and the frequency aging of each satellite clock relative to GPS time, the time and frequency offset of each monitor station clock relative to GPS time, the tropospheric residual biases for all monitor stations, and the three polar wander states. In addition to the current estimates, both the satellite orbits and the satellite clock states are predicted into the future for broadcast in the navigation message.

METHOD OF ESTABLISHING GPS AND SV TIME

- SV time is estimated by the MCS and is occasionally reset by adjusting the Z-count in the SV
- GPS time is a weighted average of the MS clocks and some SV clocks (composite clock)
- GPS time is steered to UTC(USNO) using GPS time measurements made by USNO
 - GPS time must be maintained within 1 μ s of UTC(USNO)
 - USNO provides 1 smoothed estimate of the GPS time offset each day
 - Automatic steering has been in effect since 7 January 1991 using an algorithm that changes the frequency aging of GPS time by 0 or $\pm 2 \times 10^{-19}$ s/s²
- GPS-UTC(USNO) is estimated and forecast
 - Total time transfer error budget is 110 ns

MCS COMPUTATIONS

- KALMAN FILTER ESTIMATES
 - Satellite ephemeris (6 orbital elements)
 - Solar pressure constants (2/SV)
 - SV clock bias, frequency offset, and frequency aging relative to GPS time
 - MS clock bias and frequency offset
 - tropospheric residual biases for all MS
 - polar wander (3 states)
- PREDICTED STATES FOR NAVIGATION DATA
 - Satellite ephemeris (6 orbital elements)
 - SV clock bias, frequency offset, and frequency aging relative to GPS time

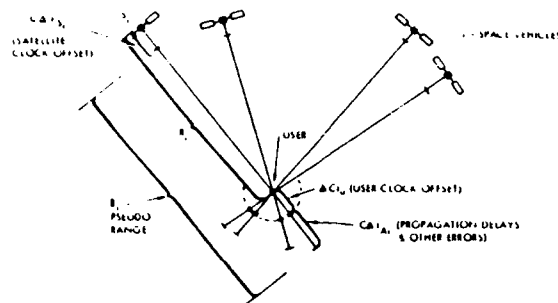
THE NAVIGATION PROBLEM – USER WITH UNKNOWN POSITION AND TIME

The next five slides summarize how GPS data is used by either an end user or the system itself to solve a navigation problem. The end user receives the GPS signal containing the information on the satellite clock times and the locations of the satellites. He computes his location, his velocity, and the time of his clock. The MCS receives the GPS signals through the monitor stations at their accurately known positions with high precision clocks and computes the positions and times of the satellites and their clocks and the times of the monitor station clocks. The computations are very similar and are based on the same concept of pseudorange. The navigation problem for the end user is solved by making ranging measurements to four satellites and solving four simultaneous linear equations for the three user position coordinates plus the time of the user clock.

GPS PSEUDORANGE MEASUREMENT

The satellite transmits a timing marker in the form of a pseudorandom noise code. The user receiver receives the signal from the satellite and correlates it against the same code generated in the user receiver. The receiver changes the delay between the local clock and the receiver generated time code until the correlation peak between the receiver generated code and the satellite code is obtained. The required delay is the offset between the satellite clock time and the user clock time.

THE NAVIGATION PROBLEM — USER WITH UNKNOWN POSITION AND TIME



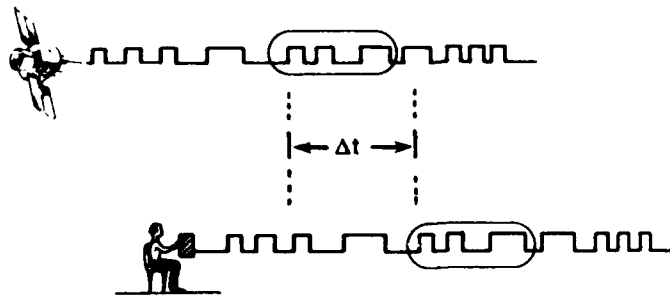
R_i = geometric range

Δt_{svi} = SVi time offset from GPS time

Δt_u = user clock offset from GPS time

Δt_{u_i} = deterministic signal delay to SVi

GPS PSEUDORANGE MEASUREMENT



GPS MEASUREMENTS

The pseudorange is the user clock reception time minus the satellite clock transmission time multiplied by the speed of light plus some corruption from noise in the process. Thus, the pseudorange is just the speed of light multiplied by the time delay measured in the GPS receiver by the correlation process. To solve for the user position, one first relates all of the times to GPS time and then relates the coordinate time delay to the transmission delay and propagation delays. The final result is that the measured pseudorange is the sum of the geometric range, the propagation delay, and the difference between the user clock and satellite clock offsets from GPS time.

USER SOLVES 4 EQUATIONS FOR 4 UNKNOWNNS

A stationary user has four unknowns: three position coordinates in the Cartesian reference frame and the time of his clock. Explicit dependence of the pseudorange on these variables is obtained by rewriting the geometric range as the square root of the sum of the squares of the differences between the Cartesian coordinates of the satellite and the user. The user receiver calculates the position of the satellite from the orbital elements transmitted in the navigation message. The user receiver also calculates the SV clock correction from the three clock parameters in the navigation message: a_0, a_1, a_2 , the time, frequency, and frequency aging of the satellite clock relative to GPS time. The relativistic correction is also calculated by the receiver based on parameters which are downloaded in the navigation message. The propagation delay is deterministic and is evaluated using either models or measurements which are discussed in greater detail in subsequent slides.

GPS MEASUREMENTS

- The raw measurement is a pseudorange

$$PR = c(\text{user clock reception time} - \text{SV clock transmission time}) + \text{noise}$$

- Relate all times to GPS (coordinate) time

$$\text{user clock reception time} = t_R + \Delta t_u$$

$$\text{SV clock transmission time} = t_{T_i} + \Delta t_{svi}$$

- The coordinate time delay is the sum of the transmission delay over the true range and the propagation delays

$$c(t_R - t_{T_i}) = R + c\Delta t_{tr}$$

- Thus

$$PR_i = R_i + c\Delta t_{tr} + c(\Delta t_u - \Delta t_{svi}) + \text{noise}_i$$

USER SOLVES 4 EQUATIONS FOR 4 UNKNOWNNS

- the form of each equation is

$$PR_i = [(x_i - x_u)^2 + (y_i - y_u)^2 + (z_i - z_u)^2]^{1/2} + c(\Delta t_u - \Delta t_{svi}) + c\Delta t_{tr} + \text{noise}_i$$

- The SV clock correction is

$$\Delta t_{svi} = a_0 + a_1(t_{T_i} - t_0) + a_2(t_{T_i} - t_0)^2 + \Delta t_r$$

where the last term is the correction for relativity and transformation to the inertial reference frame

- $c\Delta t_{tr}$ is deterministic and is evaluated using models or measurements
 - ionosphere
 - troposphere
 - antenna offsets

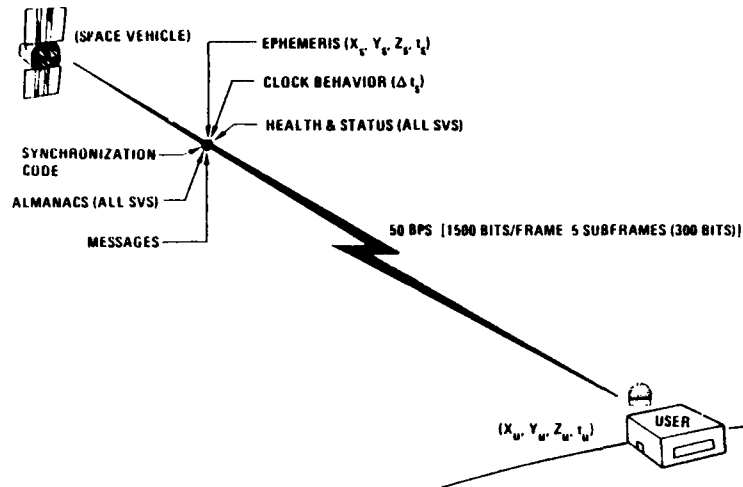
GPS NAVIGATION MESSAGE

All of the parameters required by the user for the solution of the navigation problem are downloaded from the satellite to the receiver in the navigation message which is transmitted at 50 bits per second.

RELATIVITY IN GPS

GPS uses a mixed coordinate system. The spatial coordinates of both users and satellites are referred to an earth-centered, nonrotating, nearly inertial reference frame called the ECI frame. On the other hand, the GPS time coordinate is referred to an earth-centered-earth-fixed (ECEF) reference frame. By using the Einstein synchronization process and taking into account the rotation of the reference frame, the time synchronization is made transitive using this coordinate system. For example, if three users on the surface of the earth were located at positions A, B, and C, and if A and B are synchronized and B and C are synchronized, then it follows that A and C are also synchronized. The algorithms for making relativistic corrections in the GPS have been analyzed by a variety of experts and blue ribbon panels. The results are documented and published and the conclusion is that the relativity model of GPS is correct to better than 1×10^{-14} . Since this relativity model is implemented in extremely complex computer code utilized by the MCS, the possibility remains that some small problems in the implementation of the relativity models may still exist.

GPS NAVIGATION MESSAGE



RELATIVITY IN GPS

- GPS uses a mixed coordinate system
 - coordinate time is referred to an earth centered, earth fixed, rotating reference frame (ECEF)
 - the spatial coordinates are referred to an earth centered, non rotating, nearly inertial reference frame (ECI)
- The ephemerides transmitted by the GPS satellites are ECI quantities
- Time synchronization is transitive
- The quality of the relativistic correction in GPS has been analyzed and debated throughout the development and implementation of the system
- Documented and published reports conclude that the relativity model of GPS is correct to high accuracy (10^{-14})
- This does not exclude the possibility of small difficulties in the implementation

RELATIVISTIC CORRECTIONS

The largest relativity effect is the blue shift of the satellite clock due to the difference in gravitational potential between the earth's surface and the satellite orbit. This effect of general relativity amounts to +45 microseconds per day. The next largest effect is the time dilation of the satellite clock that results from its motion. The time dilation is an effect of special relativity and amounts to approximately -7 microseconds per day. These two secular effects are primarily compensated in the manufacture of the clocks which are made to operate on the ground with a frequency of 38 microseconds per day below the 10.23 MHz nominal frequency and are shifted to close to nominal on orbit. The residual frequency offset from nominal in the operating satellite clock is absorbed in the satellite clock frequency estimate obtained by the MCS Kalman filter. From an operational point of view, it does not matter what this frequency offset is, but the MCS may choose to physically change the frequency of the on orbit clock to keep its offset from nominal frequency small. Since the GPS satellite orbits have a small eccentricity, there is a periodic relativity correction in addition to the secular frequency shifts just discussed. The correction for this relativistic time offset is computed using three parameters downloaded in the navigation message: the eccentricity of the orbit, the semi-major axis, and the eccentric anomaly.

IONOSPHERIC AND OTHER DELAYS

The L1 L2 user can correct for the ionospheric delay based on the actual measured delay to the satellite. This is possible because the ionospheric delay depends on the carrier frequency. The formula for the correction is given in ICD-GPS-200. On the other hand, the single frequency user may only correct for the nominal ionospheric delay using an ionospheric delay model. The parameters for implementing this model are transmitted in the navigation message. Since no differential delay measurement is made, the single frequency user must also correct for the group delay difference between the L1 and L2 frequencies. The tropospheric delay is much smaller but is also corrected. It is independent of transmission frequency and cannot be measured using the GPS signal. There is little uniformity in the models used for correction of the tropospheric delay. This can lead to problems for some high precision users who make differential position or timing measurements. The data collected at the monitor stations is corrected for the tropospheric delay based on measurements of local temperature, air pressure, and partial pressure of water vapor at the monitor station site.

RELATIVISTIC CORRECTIONS

- Blue shift ($+45 \mu\text{s/day}$) and time dilation ($-7 \mu\text{s/day}$) are partially removed by manufacturing the clocks with a nominal frequency offset from 10.23 MHz
- residual time invariant frequency shifts absorbed in the SV clock frequency estimate
- Correction for orbit eccentricity $\Delta t_r = F e A^{1/2} \sin[E(t)]$
where e , the eccentricity of the orbit, A , the semi-major axis, and E_k , the eccentric anomaly, are contained in the navigation message and
 $F = -4.442809305 \times 10^{-10} \text{s/m}^2$
- Effects of solar and lunar gravity cancel to 10^{-15} in vicinity of the earth

IONOSPHERIC AND OTHER DELAYS

- Two frequency ionospheric correction

$$PR = \frac{PR_{L_2} - \gamma PR_{L_1}}{1 - \gamma}$$

where $\gamma = (\nu_{L_1}/\nu_{L_2})^2$

- Single frequency user
 - Ionospheric delay model
 - The $(I_1 - I_2)$ group delay correction
- Tropospheric delay
 - Independent of frequency
 - Requires measurement of temperature, air pressure, and partial pressure of water vapor

GPS TIME TRANSFER MODES

A GPS user may synchronize his local clock directly to GPS time. This is usually called one-way time transfer or DoD time transfer since it is purely passive in operation. The user computes the difference between the measured pseudorange to a satellite and the calculated pseudorange and divides the result by the speed of light, giving his offset from GPS time. The offset of the user from UTC(USNO) is obtained in a second step by adding the GPS time corrections transmitted by the satellite to the previous result. Common view time transfer is a more precise method of synchronizing the clocks of two users. Common view is a special case of multilateration and while its use of the GPS signals is passive it requires active transmission of data between the two users. Each user obtains his time offset from GPS by the method previously described with the additional requirement that the two users always process data transmitted by the satellite at the same time. The time offset between the two users is obtained by differencing their individual time offsets from GPS. The advantage of this technique is the smaller residual time difference error resulting from various errors in the GPS system. The time error of the satellite clock and the radial ephemeris error both cancel completely. In addition, partial cancellation of the cross-track and in-track ephemeris errors and the ionospheric and tropospheric path delays is obtained.

GPS - UTC(USNO) SYNCHRONIZATION

The GPS - UTC synchronization itself is obtained via common view time transfer measurements. GPS satellites are simultaneously observed from USNO in Washington, DC and the Colorado Springs Monitor Station. Based on these measurements, USNO provides the MCS with a single smoothed time difference once each day for the time difference between GPS and UTC(USNO).

GPS TIME TRANSFER MODES

- One-way time transfer

- The pseudorange residual, *i.e.*, the measured pseudorange less the calculated pseudorange, is proportional to the offset of the user clock from GPS time

$$\Delta t_{u,GPS} = \frac{\Delta PR_i}{c} = \frac{PR_i - PR_i}{c}$$

- The offset of the user clock from UTC(USNO) is given by

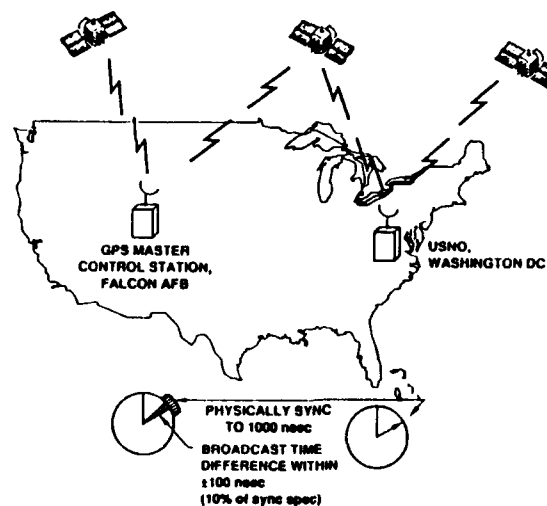
$$\Delta t_{u,UTC} = \Delta t_{u,GPS} + \Delta t_{GPS}$$

where Δt_{GPS} is UTC(USNO)-GPS and is broadcast in the navigation message

- Common-view time transfer

- Time offset between users 1 and 2 $t_{u1} - t_{u2} = \Delta t_{u1,GPS} - \Delta t_{u2,GPS}$
- Complete cancellation of SV clock error and radial ephemeris error
- Partial cancellation of cross-track and in-track ephemeris errors and ionospheric and tropospheric path delays

GPS - UTC(USNO) SYNCHRONIZATION



TIME TRANSFER ERROR BUDGET (ICD-GPS-202)

The errors in the UTC(USNO) - GPS calibration include errors in the time measurements, the predictability of GPS time, errors in the orbit determination, and a variety of user equipment errors. The total error budget specified in the ICD-GPS-202 is 110 ns (1 sigma). The following six slides give some feeling for the performance of the GPS system as a whole and its components. Satellite clock time can be observed by the user by correcting the pseudorange measurements for everything except the SV clock to GPS time offset.

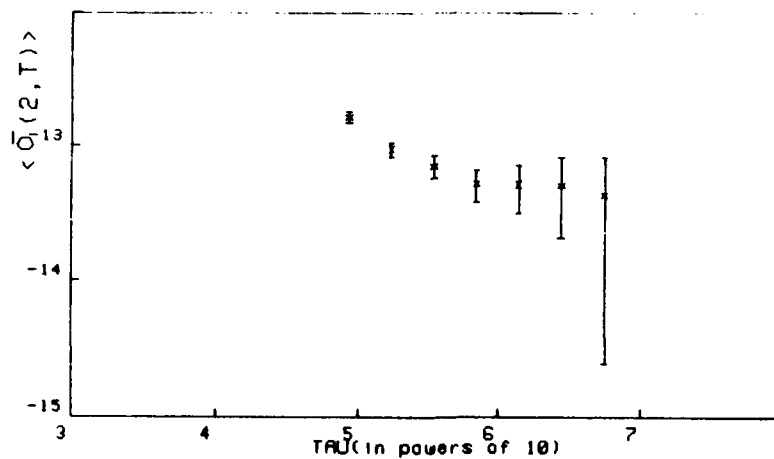
SV 13 CLOCK STABILITY

The figure shows the performance of the cesium clock on SV13, which is also designated PRN2. Clock performance is better than 1×10^{-13} for averaging times longer than approximately two days. When measured this way, the clock stability is corrupted by both ephemeris and propagation errors. Thus, the actual stability of the onboard clock might be better than the measurements indicate.

TIME TRANSFER ERROR BUDGET (ICD-GPS-202)

| COMPONENT | ERROR (ns, 1σ) |
|---|------------------------|
| Calibration prediction | |
| USNO Measurement | 70 |
| CS Measurement | 30 |
| GPS time predictability | 47 |
| Total (RSS) time difference prediction | 90 |
| Navigation message quantization | 3 |
| SV orbit | 11 |
| SV clock | 32 |
| SV group delay | 6 |
| Total (RSS) available at SV | 97 |
| Total (RSS) for downlink and user equipment | 33 |
| Total (RSS) time transfer budget | 103 |
| Total (RSS) time transfer goal | 110 |

SV 13 CLOCK STABILITY



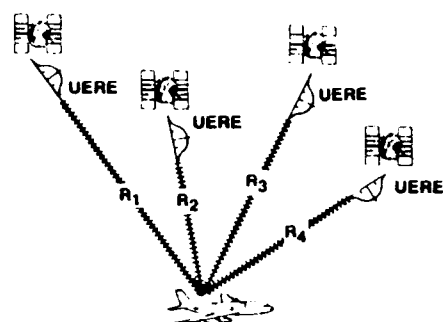
DETERMINING THE GPS NAVIGATION ACCURACY

The actual performance of the GPS system is straightforwardly described by the ranging errors between the user and the satellite. The navigation or timing errors that result from such range errors depend upon the geometry and number of satellites used for the navigation solution. For example, if a Position Dilution of Precision (PDOP) is 2.5 and the User Equivalent Range Error (UERE) is 7 meters, then the resulting navigation accuracy is 17.5 meters. Similarly, the time transfer to the user is determined from the range errors and the Time Dilution of Precision (TDOP). TDOP is the 1 sigma error in the range equivalent of the user clock offset. For example: if $UERE = 6$ meters (20ns) and $TDOP = 1.4$ (the average for a 21 S/V constellation) then user clock offset $= 1.4 \times 18 = 28ns$. This factor must be combined with residual receiver delays and other externals such as absolute position uncertainty to determine the overall one-way passive user time recovery accuracy.

NAVIGATION PERFORMANCE SUMMARY

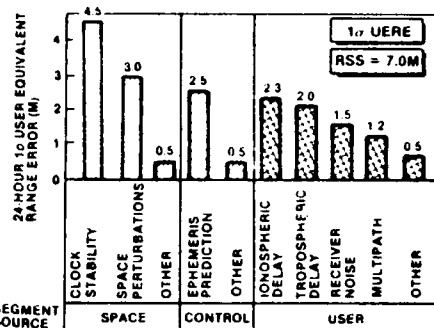
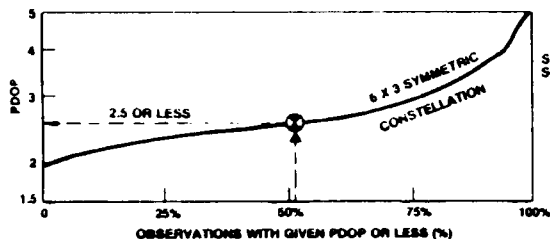
The MCS monitors system performance using a parameter called Estimated Range Deviation which is very similar to the UERE. The Estimated Range Deviations are computed for a wide variety of locations located around the world. They therefore give a globally averaged measure of range error. An individual Estimated Range Deviation is the difference between the range from the satellite to the user computed from the uploaded orbit and clock corrections and the range computed from the current Kalman estimates for the orbit and clock corrections. The graph shows the Estimated Range Deviations for the period of January 1990 to July 1991 indicating a typical performance of approximately three meters. The MCS verifies that Estimated Range Deviations are a good measure of system performance by comparing the Estimated Range Deviations for the monitor station locations with actual observed range deviations.

DETERMINING THE GPS NAVIGATION ACCURACY



THE NAVIGATION ERROR DEPENDS UPON TWO FACTORS:
1. THE RANGE ERROR TO THE SATELLITES (UERE)
2. THE GEOMETRY OF THE SITUATION (GDOP OR PDOP)

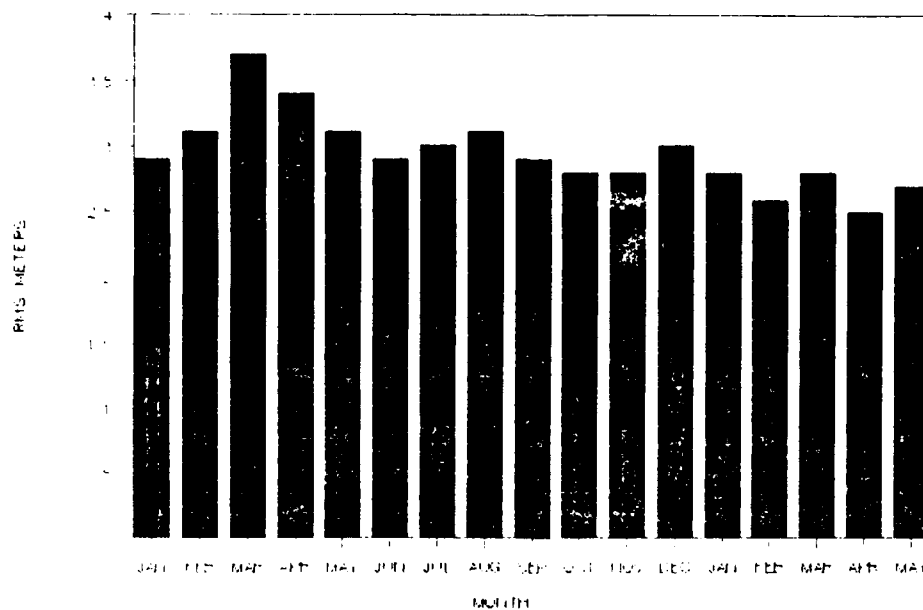
GDOP = GEOMETRICAL DILUTION OF PRECISION
PDOP = POSITION DILUTION OF PRECISION



$$\begin{aligned} \text{NAV ACCURACY} &= \text{GDOP} \times \text{UERE} \\ &= 2.5 \times 7.0\text{M} \\ &= 17\text{M} \end{aligned}$$

NAVIGATION PERFORMANCE SUMMARY

ESTIMATED RANGE DEVIATIONS — JANUARY 1990 TO JULY 1991



COPY AVAILABLE TO DTIC DOES NOT PERMIT FULLY LEGIBLE REPRODUCTION

GPS - UTC TIME TRANSFER PERFORMANCE

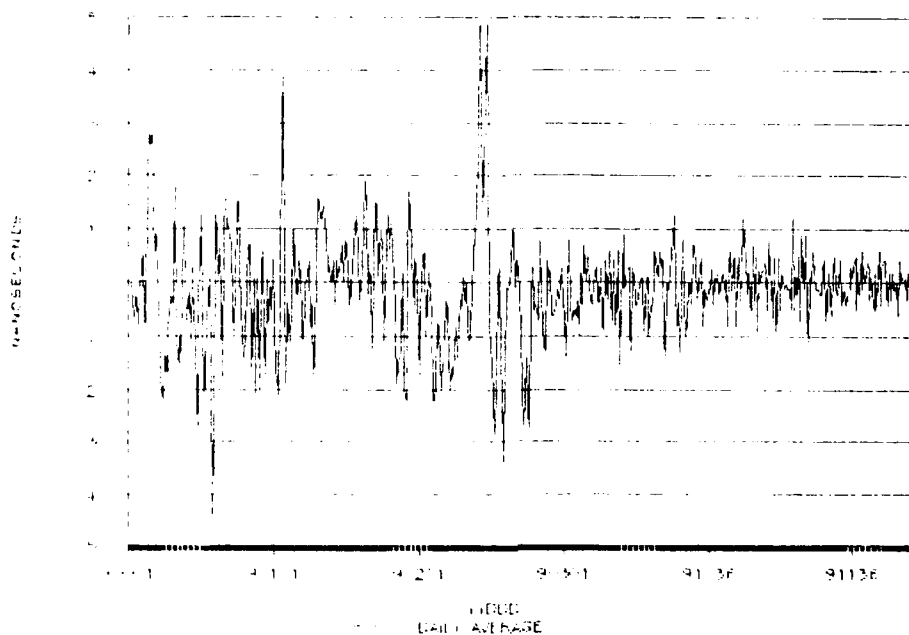
The time transfer performance of GPS has improved steadily as upgrades have been implemented within the GPS system. The graph shows time transfer performance from January 1990 to July 1991. Automatic steering of GPS was implemented throughout this whole period. The peak deviations are seen to be less than 50 ns throughout the entire period.

GPS TIME STABILITY

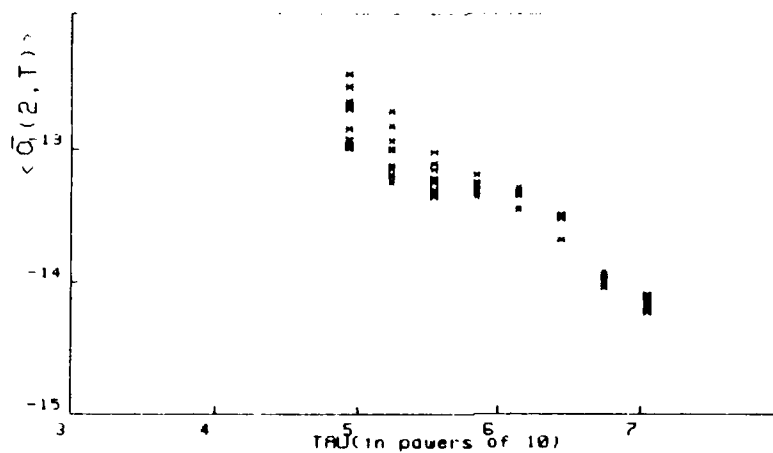
This figure shows the stability of GPS time itself as realized via one-way time transfer measurements using UTC(USNO) as a reference. Since GPS time is steered to UTC(USNO), one expects the stability to improve approximately inversely with averaging time. Although this general trend is observed, a feature may also be observed in the stability for an averaging time of approximately 10 days. The most likely explanation for the feature is the control algorithm used to steer GPS time to UTC(USNO). There are several points plotted at each averaging time. Each of these points represents the GPS time stability realized through a single satellite. There is quite a large spread in the capability of accessing GPS time depending upon the performance of the individual satellite clock.

GPS - UTC TIME TRANSFER PERFORMANCE

JANUARY 1990 TO JULY 1991



GPS TIME STABILITY



COPY AVAILABLE TO DTIC DOES NOT PERMIT FULLY LEGIBLE REPRODUCTION

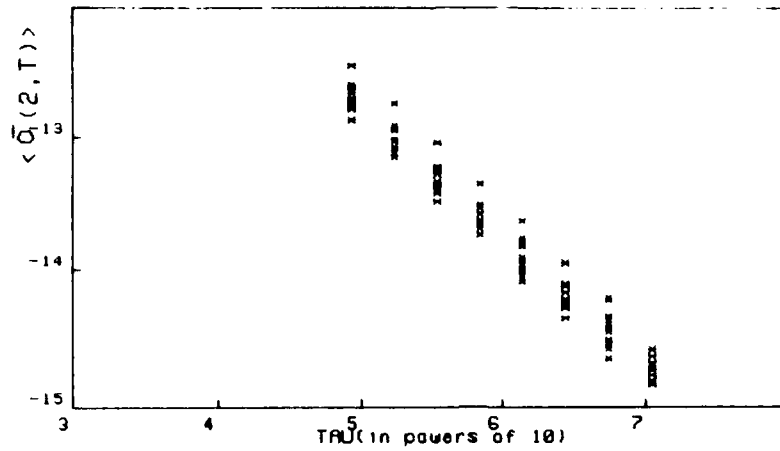
UTC ERROR

This figure shows the frequency stability of UTC realized from the GPS system compared to UTC(USNO) to which it is steered. The stability improves as one over the averaging time and there is no evidence of any feature. These data lend further credence to the thought that the feature observed on the previous figure results from the closed loop servo and not from the quality of the steering information.

DENIAL OF FULL SYSTEM ACCURACY

GPS signals transmitted from the Block 2 satellites are intentionally degraded in order to reduce the performance of the system to its original specifications for the non-authorized user. Both the satellite clock time and the ephemeris may be affected. This denial of full system capability for the standard positioning service user is called Selective Availability (SA). The term "anti-spoofing" (AS) refers to the encryption of the P code to produce Y code. Whereas the P code is a linear code available to all users, the Y code is nonlinear and can only be utilized by authorized users having the appropriate key. Authorized users with access to keyed receivers obtain the full system performance independent of both Selective Availability and Anti-Spoofing.

UTC ERROR



DENIAL OF FULL SYSTEM ACCURACY

- Selective Availability (SA)
 - Degradation of GPS to original specifications for the non-military user
 - Affects transmitted ephemeris and satellite clock
 - Block I satellites not affected
 - Block II satellites affected as they become operational
- Anti-Spoofing (AS)
 - Encryption of P-code

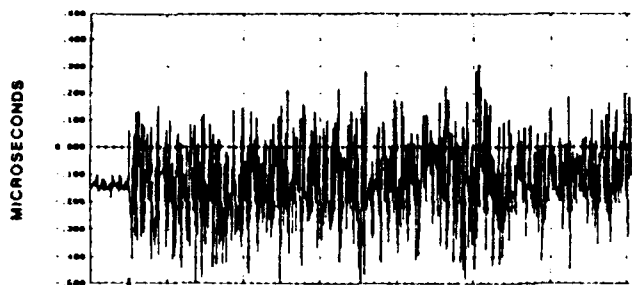
TIME TRANSFER WITH AND WITHOUT SA

This graph shows the effect of Selective Availability on SPS users. When Selective Availability was turned on, on 25 March 1990, time transfer errors were observed to increase dramatically. To the extent that these errors are either SV clock errors or radial ephemeris errors, they can be removed nearly perfectly using common view techniques. The lower plot for the same period was obtained by an authorized PPS user and shows no effect of the Selective Availability.

"BOWING" EFFECT

There are some indications that there are still significant systematic errors limiting the performance of the GPS system. One such observation has been called the "bowing effect." The graph shows the GPS to Ascension Island time transfer from all of the satellites during a period of one week. Systematic variations with this somewhat time-varying characteristic are observed having a magnitude of more than 50 ns. It is believed that the bowing effect is primarily a result of orbit estimation errors. Since the systematic effects appear to be larger than the random errors, there is some danger in using probabilistic measures for describing the overall performance of the GPS system as is commonly done.

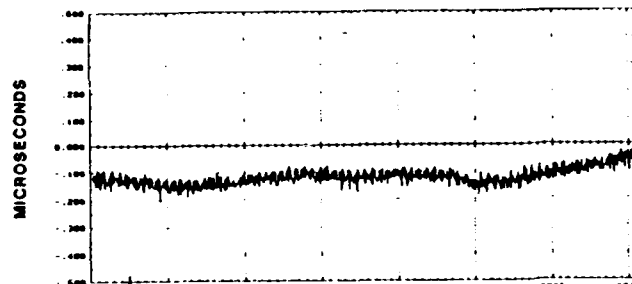
TIME TRANSFER WITH AND WITHOUT SA



USNO(MC) - GPS

STel 502 RECEIVER / SP8

BLOCK II ONLY



USNO(MC) - GPS

STel DFTMS RECEIVER / PPS

BLOCK II ONLY

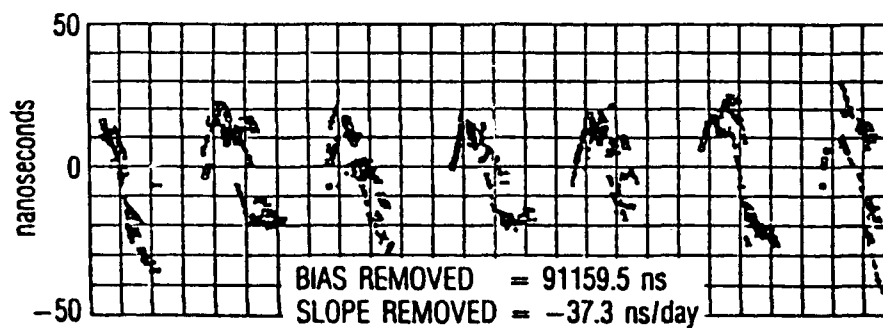
28 MARCH 90

28 MAY 90

MODIFIED JULIAN DATE (MJD)

"BOWING" EFFECT

GPS TO ASCENSION MS TIME TRANSFER



COPY AVAILABLE TO DTIC DOES NOT PERMIT FULLY LEGIBLE REPRODUCTION

SUMMARY

In summary, the GPS system is working considerably better than either the navigation or the time transfer requirements set forth in its Program Management Directive. Navigation performance is about 9 meters SEP compared to the 16 meter SEP performance requirement. Autosteering has reduced the GPS time offset from UTC to a peak of less than 30 ns compared to the requirement of 1 microsecond. One-way time transfer of 30 ns is also being achieved compared to the 110 ns requirement. Despite these generally excellent performance achievements, the user must be cautious since up to 100 ns systematic errors are sometimes observed. Finally, common view time transfer of 15 ns is routinely achieved with modest care and common view time transfer of between 1 and 2 ns is reported by various national standard laboratories.

GPS REFERENCES

"Global Positioning System", Volumes I (1980), II (1984), and III (1986), Papers published in *Navigation*, Institute of Navigation (ION), Washington, DC

"Proceedings of ION GPS-89", 1989, Institute of Navigation, Washington, DC

"Proceedings of ION GPS-90", 1990, Institute of Navigation, Washington, DC

"Proceedings of ION GPS-91", 1991, Institute of Navigation, Washington, DC

SUMMARY

- Navigation accuracy of 9 m SEP is being achieved compared to the 16 m SEP requirement
- Autosteering of GPS time to UTC(USNO) has reduced the peak offsets to 30 ns compared to a requirement of 1 μ s
- One-way time transfer of 30 ns rms is being achieved compared to the 110 ns requirement
- Up to 100 ns systematic errors are sometimes observed
- Common view time transfer of 15 ns rms is being achieved